

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:43:11 ; Search time 18.33 Seconds
(without alignments)
424.776 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 346
Sequence: 1 LRALVPHGNLQYAEIPKSEP.....ERRLDFAIRYNDWRNGNEP 346

Scoring table: OLIGO
Gapop 60.0 , Capext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A-COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B-COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A-COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B-COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS-COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	346	100.0	346	2	US-08-613-220B-4	Sequence 4, Appli
2	8	2.3	466	4	US-09-068-195-25	Sequence 25, Appl
3	7	2.0	113	4	US-08-905-223-439	Sequence 439, App
4	7	2.0	301	3	US-09-002-298-3	Sequence 3, Appli
5	7	2.0	484	4	US-08-913-578-2	Sequence 2, Appli
6	7	2.0	484	4	US-08-783-427-2	Sequence 2, Appli
7	7	2.0	549	2	US-08-676-279-59	Sequence 59, Appl
8	6	1.7	11	4	US-09-196-293-7	Sequence 7, Appli
9	6	1.7	13	5	PCT-US94-10257A-52	Sequence 52, Appl
10	6	1.7	15	1	US-08-378-761A-48	Sequence 48, Appl
11	6	1.7	15	1	US-08-485-286-48	Sequence 48, Appl
12	6	1.7	15	6	5248606-29	Patent No. 5248606
13	6	1.7	15	6	US-08-786-748A-128	Sequence 128, App
14	6	1.7	18	1	US-08-786-748A-132	Sequence 132, App
15	6	1.7	18	1	US-08-786-748A-134	Sequence 134, App
16	6	1.7	18	2	US-08-932-682-128	Sequence 128, App
17	6	1.7	18	2	US-08-932-682-132	Sequence 132, App
18	6	1.7	18	2	US-08-932-682-134	Sequence 134, App
19	6	1.7	20	5	PCT-US94-10257A-80	Sequence 80, Appl
20	6	1.7	21	1	US-08-786-748A-126	Sequence 126, App
21	6	1.7	21	1	US-08-786-748A-127	Sequence 127, App
22	6	1.7	21	1	US-08-786-748A-127	Sequence 127, App
23	6	1.7	21	1	US-08-786-748A-127	Sequence 127, App
24	6	1.7	21	1	US-08-786-748A-135	Sequence 135, App
25	6	1.7	21	1	US-08-786-748A-137	Sequence 137, App
26	6	1.7	21	2	US-08-932-682-123	Sequence 123, App
27	6	1.7	21	2	US-08-932-682-126	Sequence 126, App

Sequence 127, App
Sequence 135, App
Sequence 137, App
Sequence 17, Appl
Sequence 26, Appl
Sequence 5, Appl
Sequence 16, Appl
Patent No. 5242821
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 45, Appl
Sequence 2, Appl
Sequence 25, Appl

28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45

6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6

1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7 1.7

21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21 21

US-08-932-682-127
US-08-932-682-135
US-08-932-682-137
US-08-988-856B-17
US-08-461-990B-26
US-09-048-889-5
US-08-960-022-16
5242821-14
US-08-482-882-45
US-08-483-389-45
US-08-487-113D-45
US-08-473-503-45
US-08-483-932-45
US-08-720-420A-45
US-08-714-017-45
US-08-475-680-45
US-07-929-580B-2
US-08-609-049A-25

ALIGNMENTS

RESULT 1
US-08-613-220B-4
; Sequence 4, Application US/08613220B
; Patent No. 5958751
; GENERAL INFORMATION:
; APPLICANT: Murphy, Dennis
; TITLE OF INVENTION: ALPHA-GALACTOSIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,220B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-68-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-613-220B-4

Query Match 100.0% Score 346; DB 2; Length 346;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LRALVPHGNLOYAEIPKSEPKVIEKAYIPVIETLIKKEEPFGLNITGYTLKFLPKDIIIVK 60
Db 1 LRALVPHGNLOYAEIPKSEPKVIEKAYIPVIETLIKKEEPFGLNITGYTLKFLPKDIIIVK 60
QY 61 GGASDLIEIGTSYTAIPLPLSLRVEAQVORDRVKEELFEVSPKGFWLPDLADPIIPA 120
Db 61 GGASDLIEIGTSYTAIPLPLSLRVEAQVORDRVKEELFEVSPKGFWLPDLADPIIPA 120
QY 121 ILKDNQYELFADEAMLFSAHLSAIPKIPKPLHLIKAKRERFRYISVLLRLRLKAIK 180
Db 121 ILKDNQYELFADEAMLFSAHLSAIPKIPKPLHLIKAKRERFRYISVLLRLRLKAIK 180
QY 181 LVFEGKVTLLKVDIEAVPVVAVNAVAVMLIGRLPLMNPKNKVASWIEDKNILLYGTDIEFI 240
Db 181 LVFEGKVTLLKVDIEAVPVVAVNAVAVMLIGRLPLMNPKNKVASWIEDKNILLYGTDIEFI 240
QY 241 GYDIAGRMSVGLLEVIDELNSLCPSELKHSRGRELYLRTSSWADKSLRIWREDEGNAR 300
Db 241 GYDIAGRMSVGLLEVIDELNSLCPSELKHSRGRELYLRTSSWADKSLRIWREDEGNAR 300
QY 301 LNLVNMRCGLAPLAENSADARGWPLPERRLDFAFRAYINDWRNGNPP 346
Db 301 LNLVNMRCGLAPLAENSADARGWPLPERRLDFAFRAYINDWRNGNPP 346

RESULT 2

US-09-068-195-25
; Sequence 25, Application US/09068195B
; Patent No. 6140078
; GENERAL INFORMATION:
; APPLICANT: Sanders, Jan W.
; APPLICANT: Ledebor, Adrianus M.
; APPLICANT: Venema, Gerard
; APPLICANT: Kok, Jan

; TITLE OF INVENTION: Salt-Inducible Promoter Derivable from a Lactic Acid
; Bacterium, and Its Use in a Lactic Acid Bacterium for
; TITLE OF INVENTION: Production of Desired Protein
; FILE REFERENCE: Sanders-60113/0252227
; CURRENT APPLICATION NUMBER: US/09/068,195B
; EARLIER FILING DATE: 1998-07-29
; EARLIER FILING DATE: 1997-08-20
; EARLIER FILING DATE: 1997-03-13
; EARLIER FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 466
; TYPE: PRT
; ORGANISM: L. lactis MG1363

US-09-068-195-25

Query Match 2.3%; Score 8; DB 4; Length 466;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 DEGNARLN 302
Db 52 DEGNARLN 59

RESULT 3

US-08-905-223-439
; Sequence 439, Application US/08905223
; Patent No. 622029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duellert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS

; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 439:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: -41...-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 7.4
; OTHER INFORMATION: seq ILPLFGGIGVGFG/LF
; US-08-905-223-439

Query Match 2.0%; Score 7; DB 4; Length 113;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 TAILPLL 82
Db 27 TAILPLL 33

RESULT 4

US-09-002-298-3
; Sequence 3, Application US/09002298
; Patent No. 6046001
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Tang, Y. Tom
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: HUMAN FATTY ACID BETA-OXIDATION ENZYMES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/002,298
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0453 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 301 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: OVARUT02
CLONE: 2595635
US-09-002-298-3

Query Match 2.0%; Score 7; DB 3; Length 301;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 273 SGRELYL 279
DB 263 SGRELYL 269

RESULT 5
US-08-913-578-2
Sequence 2, Application US/08913578
Patent No. 6218159
GENERAL INFORMATION:
APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6218159el trna synthetase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,578
FILING DATE: 17-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9601069.9
FILING DATE: 19-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31352
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-913-578-2

Query Match 2.0%; Score 7; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LPHLIKA 158
DB 342 LPHLIKA 348

RESULT 6
US-08-785-427-2
Sequence 2, Application US/08785427
Patent No. 6238900
GENERAL INFORMATION:
APPLICANT: Hodgson, John
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 6238900el trna synthetase
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,427
FILING DATE: 17-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 9601069.9
FILING DATE: 19-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Gimmi, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P31352
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 484 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-785-427-2

Query Match 2.0%; Score 7; DB 4; Length 484;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 LPHLIKA 158
DB 342 LPHLIKA 348

RESULT 7
US-08-676-279-59
; Sequence 59, Application US/08676279
; Patent No. 5869247
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MACROPHAGE NUCLEOTIDE SEQUENCE
; NUMBER OF SEQUENCES: 63
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,279
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9422021.7
; FILING DATE: 31-OCT-1994
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 549 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-676-279-59

Query Match 2.0%; Score 7; DB 2; Length 549;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 39 PFGLNIT 45
DB 129 PFGLNIT 135

RESULT 8
US-09-196-293-7
; Sequence 7, Application US/09196293
; Patent No. 6183755
; GENERAL INFORMATION:
; APPLICANT: Fuchs, Renate
; APPLICANT: Motz, Manfred
; APPLICANT: Soutscheck, Erwin
; APPLICANT: Wilske, Bettina
; APPLICANT: Preac-Mursic, Vera
; TITLE OF INVENTION: Active proteins from Borrelia
; FILE REFERENCE: 738.001052
; CURRENT APPLICATION NUMBER: US/09/196,293
; EARLIER FILING DATE: 1998-11-19
; EARLIER FILING DATE: 1994-03-10
; EARLIER FILING DATE: 1992-06-19
; EARLIER FILING DATE: 1990-12-21
; EARLIER FILING DATE: 1989-12-22
; EARLIER FILING DATE: 1989-12-22
; EARLIER FILING DATE: 1990-06-13
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 11
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi

US-09-196-293-7

Query Match 1.7%; Score 6; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 250 SVEGLL 255
DB 5 SVEGLL 10

RESULT 9
US-08-209-603E-7
; Sequence 7, Application US/08209603E
; Patent No. 6248538
; GENERAL INFORMATION:
; APPLICANT: FUCHS, RENATE
; APPLICANT: WILSKE, BETTINA
; APPLICANT: PREAC-MURSIC, VERA
; APPLICANT: MOTZ, MANFRED
; APPLICANT: SOUTSCHECK, ERWIN
; TITLE OF INVENTION: IMMUNOLOGICALLY ACTIVE PROTEINS
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROOKS HAIDT HAFNER & DELAHUNTY
; STREET: 99 PARK AVENUE
; CITY: NEW YORK
; STATE: NY
; COUNTRY: USA
; ZIP: 10016
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" FLOPPY DISC
; COMPUTER: AT&T - IBM COMPATIBLE
; OPERATING SYSTEM: MS-DOS Version 6.2
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/209,603E
; FILING DATE: 10-MAR-1994
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP90/02282
; FILING DATE: 21-DEC-1990
; APPLICATION NUMBER: US 07/862,535
; FILING DATE: 19-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINSON, WILLIAM R.
; REGISTRATION NUMBER: 27,224
; REFERENCE/DOCKET NUMBER: LKR-9217-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 697-3355
; TELEFAX: (212) 557-5635
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE:
; DESCRIPTION: PROTEIN
; HYPOTHETICAL: N/A
; ANTI-SENSE: N/A
; FRAGMENT TYPE: N/A
; ORIGINAL SOURCE:
; ORGANISM: B. BURGDOFFERI
; IMMEDIATE SOURCE:
; LIBRARY: DSM 5662
; POSITION IN GENOME: N/A
; FEATURE:
; IDENTIFICATION METHOD: amino acid analysis
; PUBLICATION INFORMATION: N/A
US-08-209-603E-7

Query Match 1.7%; Score 6; DB 4; Length 11;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 SVEGLL 255
|||||
Db 5 SVEGLL 10

RESULT 10
PCT-US94-10257A-52
; Sequence 52, Application PC/TUS9410257A
; GENERAL INFORMATION:
; APPLICANT: IMMULOGIC PHARMACEUTICAL CORPORATION
; APPLICANT: BRIGITTE DEVAUX
; APPLICANT: JONATHAN B. ROTHBARD
; APPLICANT: DAWN SMILEK
; TITLE OF INVENTION: USES OF MYELIN OLIGODENDROCYTE
; TITLE OF INVENTION: GLYCOPROTEIN AND PEPTIDE PORTIONS THEREOF IN PROTOCOLS
; TITLE OF INVENTION: RELATED TO AUTOIMMUNE DISEASE
; NUMBER OF SEQUENCES: 95
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: IMMULOGIC PHARMACEUTICAL CORPORATION
; STREET: 610 LINCOLN STREET
; CITY: WALTHAM
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02145
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/10257A
; FILING DATE: 1 SEPTEMBER 1994
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/116,824
; FILING DATE: 03-SEP-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: ANNE I CRAIG
; REGISTRATION NUMBER: 32,976
; REFERENCE/DOCKET NUMBER: 071.1 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 466-6000
; TELEFAX: (617) 466-6040
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: internal
PCT-US94-10257A-52

Query Match 1.7%; Score 6; DB 5; Length 13;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 184 EGKVTL 189
|||||
Db 5 EGKVTL 10

RESULT 11
US-08-378-761A-48
; Sequence 48, Application US/08378761A
; Patent No. 5635384
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,761A
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-378-761A-48

Query Match 1.7%; Score 6; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RFRYS 168
|||||
Db 10 RFRYS 15

RESULT 12
US-08-485-286-48
; Sequence 48, Application US/08485286
; Patent No. 5646026
; Patent No. 5646026 5646119
; GENERAL INFORMATION:
; APPLICANT: WALSH, TERENCE A
; APPLICANT: HEY, TIMOTHY D
; TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE
; TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ANDREA T. BORUCKI
; STREET: 9330 ZIONSVILLE ROAD
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,286
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/378761
; FILING DATE: 26-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BORUCKI, ANDREA T
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 38272B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 337-4846
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-286-48

Query Match 1.7%; Score 6; DB 1; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RFRYS 168
| | | | |
Db 10 RFRYS 15

RESULT 13
5248606-29
; Patent No. 5248606
; ALICE E.R.
; TITLE OF INVENTION: DNA ENCODING INACTIVE PRECURSOR AND
; ACTIVE FORMS OF MAIZE RIBOSOME INACTIVATING
; NUMBER OF SEQUENCES: 49
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/535,636
; FILING DATE: 11-JUN-1990
; SEQ ID NO: 29:
; LENGTH: 15
5248606-29

Query Match 1.7%; Score 6; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 163 RFRYS 168
| | | | |
Db 10 RFRYS 15

RESULT 14
US-08-786-748A-128
; Sequence 128, Application US/08/86748A
; Patent No. 5714577
; GENERAL INFORMATION:
; APPLICANT: Ronald, Montelaro C.
; APPLICANT: Tencza, Sarah B.
; APPLICANT: Mietzner, Timothy A.
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA

; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,748A
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/010,634
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Rochelle K. Seide
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: AP30421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-408-2500
; TELEFAX: 212-765-2519
; INFORMATION FOR SEQ ID NO: 128:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 18 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5714577e
US-08-786-748A-128

Query Match 1.7%; Score 6; DB 1; Length 18;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GLLEVI 258
| | | | |
Db 3 GLLEVI 8

RESULT 15
US-08-786-748A-132
; Sequence 132, Application US/08/86748A
; Patent No. 5714577
; GENERAL INFORMATION:
; APPLICANT: Ronald, Montelaro C.
; APPLICANT: Tencza, Sarah B.
; APPLICANT: Mietzner, Timothy A.
; TITLE OF INVENTION: NOVEL ANTIMICROBIAL PEPTIDES
; NUMBER OF SEQUENCES: 169
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/786,748A
; FILING DATE: 24-JAN-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/010,634
; FILING DATE: 26-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Rochelle K. Seide
; REGISTRATION NUMBER: 32,300
; REFERENCE/DOCKET NUMBER: AP30421
; TELECOMMUNICATION INFORMATION:

us-09-407-806a-4.rai

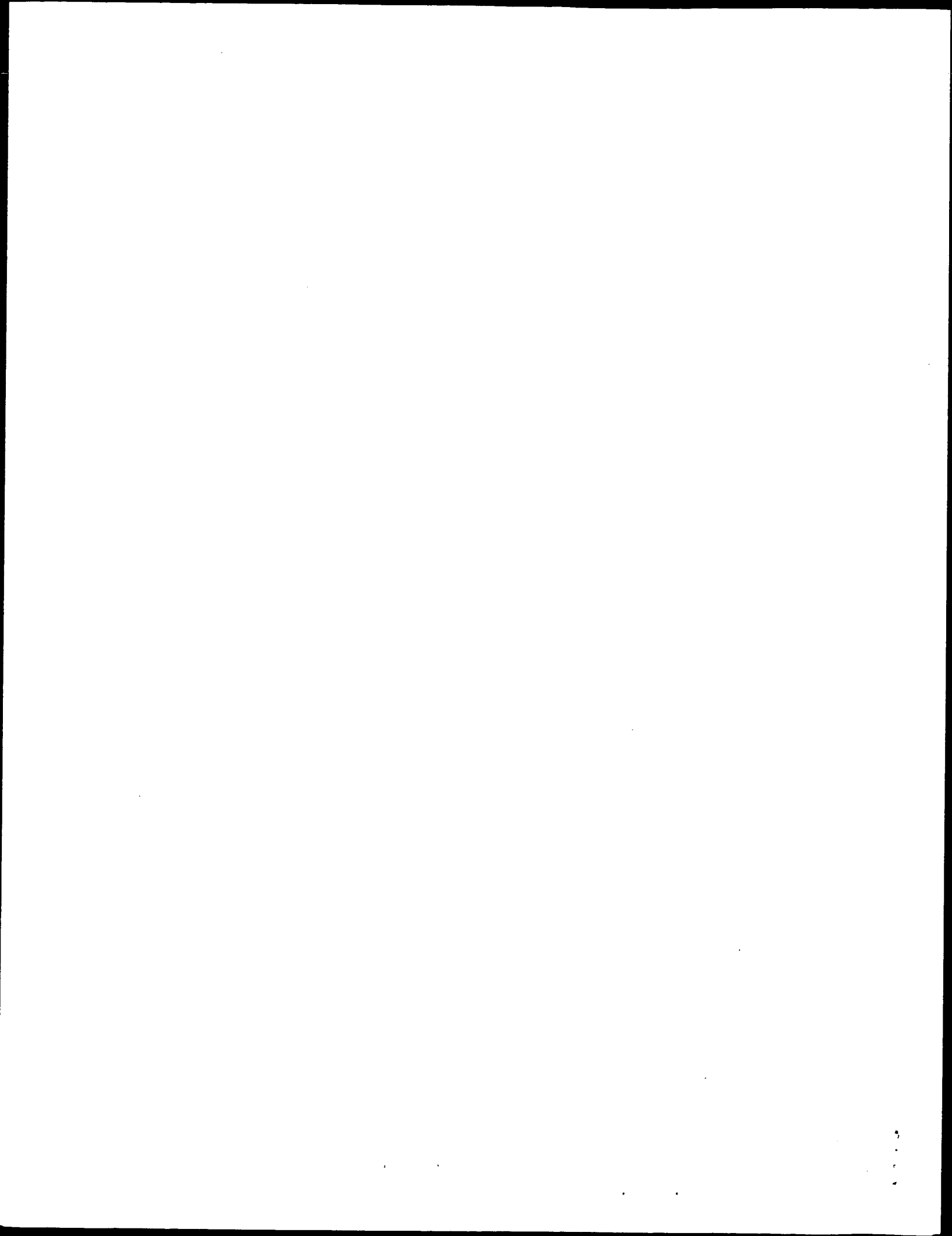
Thu Apr 4 16:34:53 2002

TELEPHONE: 212-408-2500
 TELEFAX: 212-765-2519
 INFORMATION FOR SEQ ID NO: 132:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 18 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: No. 5714577e
 US-08-786-748A-132

Query Match 1.78; Score 6; DB 1; Length 18;
 Best Local Similarity 100.0%; Pred. No. 37;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 253 GLEVI 258
 |||||
 Db 3 GLEVI 8

Search completed: April 4, 2002, 14:46:01
 Job time: 170 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:43:52 ; Search time 21.66 seconds
(without alignments)
1216.823 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 346
Sequence: 1 LRALVFHGNLQYAEIPKSEP.....ERRLDFAFRAINDWRNGNEP 346

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size : 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_68:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	12	3.5	364	E71144	hypothetical prote
2	8	2.3	218	A64445	probable phospho
3	8	2.3	243	S75330	hypothetical prote
4	8	2.3	272	H75219	hypothetical prote
5	8	2.3	272	A71203	hypothetical prote
6	8	2.3	466	B86786	glutamate decarbox
7	8	2.3	492	T01584	cellulase (EC 3.2.
8	8	2.3	529	A47114	phosphoprotein pho
9	8	2.3	619	F81532	conserved hypothet
10	8	2.3	619	F72006	CT858 hypothetical
11	8	2.3	619	E86617	CT858 hypothetical
12	8	2.3	878	S47332	DNA topoisomerase
13	7	2.0	128	D86809	ferric uptake regu
14	7	2.0	147	RPBPF5	immunity repressor
15	7	2.0	175	I G69518	isochorismatase (e
16	7	2.0	178	I 140072	translation initia
17	7	2.0	195	S34993	nitrite reductase
18	7	2.0	231	H64466	hypothetical prote
19	7	2.0	248	C83865	hypothetical prote
20	7	2.0	255	B70982	probable nei prote
21	7	2.0	267	T A75131	hypothetical prote
22	7	2.0	277	T 71456	hypothetical prote
23	7	2.0	302	E72402	phospho-N-acetylmu
24	7	2.0	304	G69601	integrase/recombin
25	7	2.0	318	D82742	dihydroxydipicolin
26	7	2.0	325	T 17307	hypothetical prote
27	7	2.0	326	T A83273	fatty acid biosynt
28	7	2.0	330	1 YWBS	tryptophan--trNA l
29	7	2.0	333	2 H75637	probable transposa

ALIGNMENTS

RESULT 1

E71144
hypothetical protein PH0368 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000

C:Accession: E71144

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hino, Y.; Yamamoto, S.; S. M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Og

DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: E71144

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-364 <KAW>

A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29442.1; PID:d1030385; PID:g3

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenB

C:Genetics:

A:Gene: PH0368

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0368

Query Match 3.5%; Score 12; DB 2; Length 364;
Best Local Similarity 100.0%; Pred. No. 0.00071;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 FHGNLQYAEIPK 17

DB 6 FHGNLQYAEIPK 17

RESULT 2

A64445
probable phosphoesterase (EC 3.1.1.-) MJ1162 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: A64445

R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Bla
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannas

A:Reference number: A64300; MUID:96337999

A:Accession: A64445

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-218 <BUL>

A:Cross-references: GB:U67558; GB:L77117; NID:g1591786; PIDN:AAB99164.1; PID:g159178

C:Genetics:

A:Map position: REV1103956-1103300

C:Superfamily: probable phosphoesterase MTH1882; phosphoesterase core homology
 C:Keywords: hydrolase
 F:2-67/Domain: phosphoesterase core homology <PEC>

Query Match 2.3%; Score 8; DB 2; Length 218;
 Best Local Similarity 100.0%; Pred. No. 4.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 EVIDELNS 263

|||||||

Db 70 EVIDELNS 77

RESULT 3

S75330

hypothetical protein sll1570 - *Synechocystis* sp. (strain PCC 6803)

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997

C:Accession: S75330 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S75330

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-243 <KAW>

A:Cross-references: EMBL:D90904; GB:AB001339; NID:gl652225; PIDN:BAA17244.1; PID:gl65232

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

C:Superfamily: *Synechocystis* hypothetical protein sll1570

Query Match 2.3%; Score 8; DB 2; Length 243;

Best Local Similarity 100.0%; Pred. No. 5.3;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PKKVASWI 225

|||||||

Db 16 PKKVASWI 23

RESULT 4

H75219

hypothetical protein PAB2165 - *Pyrococcus abyssi* (strain Orsay)

C:Species: *Pyrococcus abyssi*

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000

C:Accession: H75219

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: H75219

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-272 <KAW>

A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49207.1; PID:g545771

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB2165

C:Superfamily: *Pyrococcus abyssi* hypothetical protein PAB2165

Query Match

Best Local Similarity 2.3%; Score 8; DB 2; Length 272;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LLEVIDEL 261

Db 173 LLEVIDEL 180

|||||||

RESULT 5

A71203

hypothetical protein PH1894 - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 28-Jul-2000

C:Accession: A71203

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: A71203

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-272 <KAW>

A:Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31016.1; PID:g3258333

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH1894

C:Superfamily: *Pyrococcus abyssi* hypothetical protein PAB2165

Query Match

Best Local Similarity 2.3%; Score 8; DB 2; Length 272;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LLEVIDEL 261

|||||||

Db 173 LLEVIDEL 180

RESULT 6

B86786

glutamate decarboxylase (EC 4.1.1.15) [imported] - *Lactococcus lactis* subsp. *lactis*

C:Species: *Lactococcus lactis* subsp. *lactis*

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Mar-2001

C:Accession: B86786

R:Boletín, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissbach, J.; Eh

Genome Res. in press, 2001

A:Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: B86786

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-466 <STO>

A:Cross-references: GB:AE005176; NID:gl2724267; PIDN:AAK05388.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: gadB

C:Superfamily: *Escherichia coli* glutamate decarboxylase

C:Keywords: carbon-carbon lyase; carboxy-lyase

Query Match

Best Local Similarity 2.3%; Score 8; DB 2; Length 466;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 DEGNARLN 302

|||||||

Db 52 DEGNARLN 59

RESULT 7

T01584

cellulase (EC 3.2.1.4) F16B22.6 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 16-Feb-2001

C:Accession: T01584; B84880

Thu Apr 4 16:34:54 2002

us-09-407-806a-4.rpr

R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, submitted to the EMBL Data Library, July 1998

A:Description: Arabidopsis thaliana chromosome II BAC F16B22 genomic sequence.

A:Reference number: Z14284

A:Accession: T01584

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-492 <ROU>

A:Cross-references: EMBL:AC003672; NID:g3341671; PID:g3341677

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.W.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: B84880

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-492 <STO>

A:Cross-references: GB:AE002093; NID:g3341677; PID:AAC27459.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g44570; F16B22.6

A:Map position: 2

A:Introns: 74/3; 292/3; 387/3

C:Superfamily: Arabidopsis membrane-anchored cellulase KOR

C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 2.3%; Score 8; DB 2; Length 492;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 KLVEGKV 187

Db 292 KLVEGKV 299

RESULT 8

A47114

phosphoprotein phosphatase (EC 3.1.3.16) 2A regulatory chain PR72 - human

C:Species: Homo sapiens (man)

C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 08-Oct-1999

R:Accession: A47114

R:Hendrix, P.; Mayer-Jaekel, R.E.; Cron, P.; Goris, J.; Hofsteenge, J.; Merlievede, W.; H

J. Biol. Chem. 268, 15267-15276, 1993

A:Title: Structure and expression of a 72-kDa regulatory subunit of protein phosphatase

A:Reference number: A47114; MUID:93315512

A:Accession: A47114

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-529 <HEN>

A:Cross-references: GB:L12146; NID:g190221; PID:AA02614.1; PID:g190222

C:Keywords: alternative splicing; phosphoric monoester hydrolase

Query Match 2.3%; Score 8; DB 2; Length 529;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 RGEAFLA 315

Db 15 RGEAFLA 22

RESULT 9

F81532

conserved hypothetical protein CP0837 [imported] - Chlamydomonas reinhardtii (strain AR39)

C:Species: Chlamydomonas reinhardtii, Chlamydia pneumoniae

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000

C:Accession: F81532

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPh and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255

A:Accession: F81532

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-619 <REA>

A:Cross-references: GB:AE002243; GB:AE002161; NID:g7189750; PID:AAF38630.1; PID:g718

A:Experimental source: strain AR39, HL cells

C:Genetics:

A:Gene: CP0837

Query Match 2.3%; Score 8; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PAILKONG 126

Db 498 PAILKONG 505

RESULT 10

F72006

CT858 hypothetical protein - Chlamydomonas reinhardtii (strain CWL029)

C:Species: Chlamydomonas reinhardtii, Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C:Accession: F72006

R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: F72006

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-619 <ARN>

A:Cross-references: GB:AE001682; GB:AE001363; NID:g4377344; PID:AA019153.1; PID:g437

A:Experimental source: strain CWL029

C:Genetics:

A:Gene: CPN1016

Query Match 2.3%; Score 8; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PAILKONG 126

Db 498 PAILKONG 505

RESULT 11

E86617

CR858 hypothetical protein [imported] - Chlamydomonas reinhardtii (strain J138)

C:Species: Chlamydomonas reinhardtii, Chlamydia pneumoniae

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001

C:Accession: E86617

R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.

Nucleic Acids Res. 28, 2311-2314, 2000

A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.

A:Reference number: A86491; MUID:20330349

A:Accession: E86617

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-619 <STO>

A:Cross-references: GB:BA000008; NID:g8979389; PID:BAA99223.1; GSPDB:GN00142

A:Experimental source: strain J138

C:Genetics:

A:Gene: CPJ1016

Query Match 2.3%; Score 8; DB 2; Length 619;
Best Local Similarity 100.0%; Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 FAILKDNG 126
|||||
Db 498 FAILKDNG 505

RESULT 12

S47332

DNA topoisomerase (ATP-hydrolyzing) (EC 5.99.1.3) - Erwinia carotovora

C:Species: Erwinia carotovora

C:Date: 23-Nov-1994 #sequence_revision 03-Aug-1995 #text_change 16-Jul-1999

C:Accession: S47332

R:Rosanas, A.

submitted to the EMBL Data Library, August 1994

A:Reference number: S47332

A:Accession: S47332

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-878 <ROS>

A:Cross-references: EMBL:X80798; NID:g525201; PIDN:CAA56777.1; PID:g525202

C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phase T4 DNA topoisomerase

C:Keywords: ATP; isomerase

F:1-238/Domain: phase T4 DNA topoisomerase (ATP-hydrolyzing) medium chain homology <T4>

Query Match

Best Local Similarity 2.3%; Score 8; DB 2; Length 878;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 DEAMLFSA 140

|||||

Db 691 DEAMLFSA 698

RESULT 13

D86809

ferric uptake regulator [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C:Species: Lactococcus lactis subsp. lactis

C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 23-Mar-2001

C:Accession: D86809

R:Botolin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. in press, 2001

A>Title: The complete genome sequence of the lactic acid bacterium.

A:Reference number: A86625

A:Accession: D86809

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-128 <STO>

A:Cross-references: GB:AE005176; NID:gl2724470; PIDN:AAK05574.1; GSPDB:GN00146

A:Experimental source: strain IL1403

C:Genetics:

A:Gene: fur

Query Match

Best Local Similarity 2.0%; Score 7; DB 2; Length 128;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GLNITGY 47

|||||

Db 113 GLNITGY 119

RESULT 14

RBPFF5

immunity repressor protein - Bacillus phage phi-105

C:Species: Bacillus phage phi-105

A:Note: host Bacillus subtilis

C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 16-Jun-2000

C:Accession: A93579; A91535; S02459; T13543; A24339; A24521

R:Dhaese, P.; Seurinck, J.; De Smet, B.; Van Montagu, M.

Nucleic Acids Res. 13, 5441-5455, 1985

A>Title: Nucleotide sequence and mutational analysis of an immunity repressor gene fr

A:Reference number: A93579; MUID:85297750

A:Accession: A93579

A:Molecule type: DNA

A:Residues: 1-147 <DHA>

A:Cross-references: GB:X02799; NID:gl5455; PIDN:CAA36567.1; PID:g579178

R:Cully, D.F.; Garro, A.J.

Gene 38, 153-164, 1985

A>Title: Nucleotide sequence of the immunity region of Bacillus subtilis bacteriophage

A:Reference number: A91535; MUID:86056972

A:Accession: A91535

A:Molecule type: DNA

A:Residues: 1-147 <CUL>

R:van Kaer, L.; Ganssems, Y.; van Montagu, M.; Dhaese, P.

EMBO J. 7, 859-866, 1988

A>Title: Interaction of the Bacillus subtilis phage phi105 repressor with operator DN

A:Reference number: S02459; MUID:88283656

A:Accession: S02459

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 4-147 <VAN>

A:Note: Part of this sequence, including the amino end of the mature protein, was con

R:Kobayashi, K.; Okamura, K.; Inoue, T.; Sato, T.; Kobayashi, Y.

submitted to the EMBL Data Library, July 1998

A:Description: Complete nucleotide sequence of Bacillus subtilis phage phi-105.

A:Reference number: Z17688

A:Accession: T13543

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 4-147 <KOB>

A:Cross-references: EMBL:AB016282; PIDN:BAA36660.1

C:Genetics:

A:Start codon: GTG

C:Superfamily: phage phi-105 immunity repressor protein

C:Keywords: DNA binding; early protein; repressor; transcription regulation

Query Match

Best Local Similarity 2.0%; Score 7; DB 1; Length 147;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 ETLIKEE 38

|||||

Db 68 ETLIKEE 74

RESULT 15

G69518

isochorismatase (entB) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: G69518

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

.: Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch

A:Reference number: A69250; MUID:98049343

A:Accession: G69518

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-175 <KLE>

A:Cross-references: GB:AE000956; GB:AE000782; NID:g2689279; PIDN:AA889104.1; PID:g264

C:Superfamily: hypothetical protein b1011

Query Match

Best Local Similarity 2.0%; Score 7; DB 1; Length 175;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 EVIDELN 262

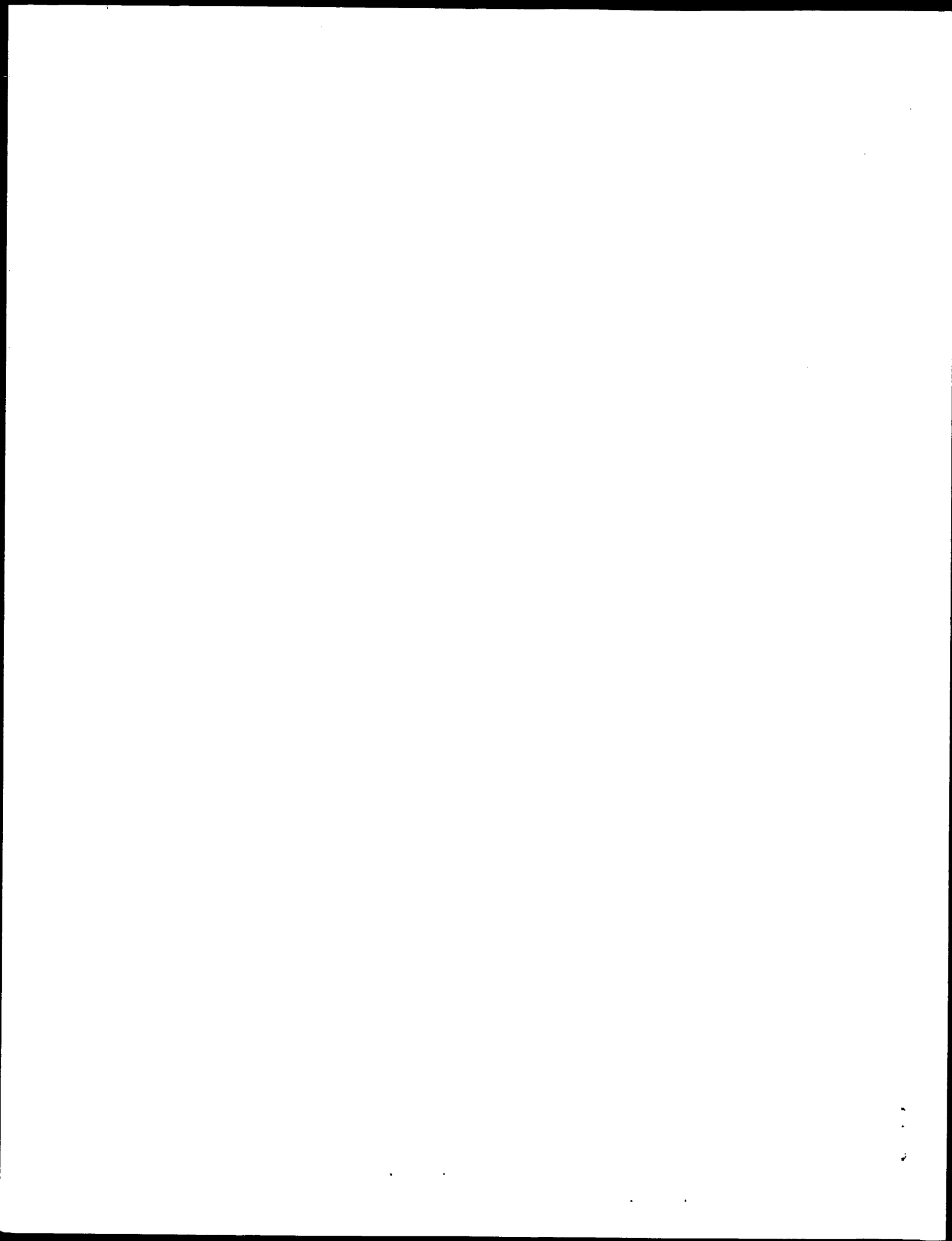
Thu Apr 4 16:34:54 2002

us-09-407-806a-4.rpr

Page 5

Db 74 EVIDELN 80
 |||||||

Search completed: April 4, 2002, 14:46:37
Job time: 165 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:46:02 ; Search time 12.23 Seconds
(without alignments)
1037.288 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 346
Sequence: 1 LRLVFGNLYAEIPKSEP.....ERRLDFAIRYNDWRNGEP 346

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 100059 seqs, 36664827 residues
Word size : 0
Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.3	619	1 YA16_CHLPN	Q926p3 chlamydia p
2	8	2.3	878	1 GYRA_ERWCA	P415l3 erwinia car
3	7	2.0	144	1 RYC_BPPH1	P06153 bacterioph
4	7	2.0	151	1 SYB2_RHIME	Q923q1 rhizobium m
5	7	2.0	178	1 IF3_BUCAP	P46243 buchnera ap
6	7	2.0	231	1 YD37_METJA	Q58733 methanococc
7	7	2.0	237	1 Y4MC_RHISN	P96902 mycobacteri
8	7	2.0	255	1 ENDB_MYCTU	P95562 rhizobium s
9	7	2.0	302	1 DAPA_XYLFA	Q9per5 xylella fas
10	7	2.0	302	1 MRAY_THEMA	Q9wy77 thermotoga
11	7	2.0	304	1 CODV_BACSU	P39776 bacillus su
12	7	2.0	321	1 LPXK_RICRI	P58187 rickettsia
13	7	2.0	325	1 LPXK_RICMO	P58186 rickettsia
14	7	2.0	330	1 SYW_BACSU	P21656 bacillus su
15	7	2.0	354	1 GBA3_USTHO	O14438 ustilago ho
16	7	2.0	358	1 PIAP_PIG	Q62640 sus scrofa
17	7	2.0	450	1 Y425_MYCPN	P75172 mycoplasma
18	7	2.0	549	1 SMF2_YEAST	P38778 saccharomyc
19	7	2.0	601	1 Y248_CHLMO	Q9p160 chlamydia m
20	7	2.0	632	1 ARPB_ECOLI	P76205 escherichia
21	7	2.0	644	1 YED0_YEAST	P39987 saccharomyc
22	7	2.0	666	1 YB9C_YEAST	P38339 saccharomyc
23	7	2.0	709	1 UVRE_MICLO	P10125 micrococcus
24	7	2.0	815	1 RPGR_HUMAN	Q92834 homo sapien
25	7	2.0	1237	1 POLA_DROME	P10394 drosophila
26	7	2.0	1320	1 PUR4_NEIMB	Q9jxk5 neisseria m
27	7	2.0	2499	1 MPRI_BOVIN	P08169 bos taurus
28	7	2.0	4684	1 PLEI_HUMAN	Q15149 homo sapien
29	7	2.0	4687	1 PLEI_RAT	P30427 rattus norv
30	6	1.7	24	1 ALKE_SALTY	P37462 salmonella
31	6	1.7	63	1 ANPI_AUSBR	P12100 austrolycic
32	6	1.7	63	1 ANP2_AUSBR	P12101 austrolycic
33	6	1.7	63	1 RL29_VIBCH	Q9knz2 vibrio chol

RESULT 1

YAL6_CHLPN	1	6	1.7	70	1	RL29_HALMA	PI0971 haloarcula
AC	Q926P3; O9JS57;	6	1.7	94	1	CH10_THERB	O60023 thermoanaer
DT	30-MAY-2000 (Rel. 39, Created)	6	1.7	105	1	YNC8_YEAST	P53967 saccharomyc
DT	20-AUG-2001 (Rel. 40, Last sequence update)	6	1.7	107	1	QACG_STASP	O87866 staphylococ
DE	20-AUG-2001 (Rel. 40, Last annotation update)	6	1.7	112	1	GLNB_PORPU	P51254 porphyra pu
DE	HYPOTHETICAL PROTEIN CPN1016/CP0837/CPJ1016.	6	1.7	118	1	Y493_HAEIN	O05023 haemophilus
GN	CPN1016 OR CP0837 OR CPJ1016.	6	1.7	125	1	RS6_PASMU	O9cln8 pasteurella
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).	6	1.7	128	1	C562_ECOLI	P00192 escherichia
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	6	1.7	128	1	RS6_THEMA	O9wz72 thermotoga
OX	NCHI_TaxID=83558;	6	1.7	129	1	EXBD_HAEDU	O51809 haemophilus
RN	[1]	6	1.7	133	1	YBB2_YEAST	P38209 saccharomyc
RP	SEQUENCE FROM N.A.	6	1.7	134	1	ANP3_RHIDE	P35753 rhigophila
RC	STRAIN=RWL029;						
RX	MEDLINE=92026606; PubMed=10192388;						
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,						
RA	Olinger L., Greenwood J., Davis R.W., Stephens R.S.;						
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";						
RL	Nat. Genet. 21:385-389(1999).						
RN	[2]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=AR39;						
RX	MEDLINE=20150255; PubMed=10684935;						
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,						
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,						
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,						
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,						
RA	Eisen J., Fraser C.M.;						
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia						
RL	pneumoniae AR39.";						
RN	[3]						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=J138;						
RX	MEDLINE=20330349; PubMed=10871362;						
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,						
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;						
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138						
RT	from Japan and CWL029 from USA.";						
RL	Nucleic Acids Res. 28:2311-2314(2000).						
CC	-!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN1016/CT058/TC0248						
CC	FAMILY.						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/						
CC	or send an email to license@isb-sib.ch).						
CC	EMBL; AE001682; AAD19153.1; -						
DR	EMBL; AE002243; AAF38630.1; -						
DR	EMBL; AP002548; BAA99223.1; -						
DR	TIGR; CP0837; -						

ALIGNMENTS

YAL6_CHLPN	STANDARD;	PRT;	619 AA.
AC	Q926P3; O9JS57;		
DT	30-MAY-2000 (Rel. 39, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	HYPOTHETICAL PROTEIN CPN1016/CP0837/CPJ1016.		
GN	CPN1016 OR CP0837 OR CPJ1016.		
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).		
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.		
OX	NCHI_TaxID=83558;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=RWL029;		
RX	MEDLINE=92026606; PubMed=10192388;		
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,		
RA	Olinger L., Greenwood J., Davis R.W., Stephens R.S.;		
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";		
RL	Nat. Genet. 21:385-389(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=AR39;		
RX	MEDLINE=20150255; PubMed=10684935;		
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,		
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,		
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,		
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,		
RA	Eisen J., Fraser C.M.;		
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia		
RL	pneumoniae AR39.";		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=J138;		
RX	MEDLINE=20330349; PubMed=10871362;		
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,		
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;		
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138		
RT	from Japan and CWL029 from USA.";		
RL	Nucleic Acids Res. 28:2311-2314(2000).		
CC	-!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN1016/CT058/TC0248		
CC	FAMILY.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	EMBL; AE001682; AAD19153.1; -		
DR	EMBL; AE002243; AAF38630.1; -		
DR	EMBL; AP002548; BAA99223.1; -		
DR	TIGR; CP0837; -		

DR InterPro: IPR000178; IF2.
 DR InterPro: IPR003581; TSPc.
 DR ProDom: PD186100; IF2; 1.
 DR SMART: SM00245; TSPc; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 619 619
 SQ SEQUENCE 619 AA; 69178 MW; 25C3B607B5936FDB CRC64;

Query Match 2.3%; Score 8; DB 1; Length 619;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 PAILKONG 126
 DB 498 PAILKONG 505

RESULT 2
 ID GYRA_ERWCA STANDARD; PRT; 878 AA.
 AC P41513;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE DNA GYRASE SUBUNIT A (EC 5.99.1.3).
 GN GYRA.

OS Erwinia carotovora.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Pectobacterium.
 OX NCBI_TaxID=554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95369719; PubMed=7642123;
 RA Rosanas A., Barbe J., Gilbert I.;
 RT "Cloning and sequencing of the gyra gene from the plant pathogen
 RT Erwinia carotovora.";
 RL Gene 161:11-14(1995).

CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
 CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
 CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
 CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
 CC -!- CATALYTIC ACTIVITY: ATP-DEPENDENT BREAKAGE, PASSAGE AND REJOINING
 CC OF DOUBLE-STRANDED DNA.
 CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
 CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
 CC ENZYME FORMS AN A2B2 TETRAMER.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X80798; CAA56777.1;
 DR InterPro: IPR002205; DNA_topoisomIV.
 DR Pfam: PF00521; DNA_topoisomIV; 1.
 DR SMART: SM00434; TOP4c; 1.
 KW Topoisomerase; Isomerase; DNA-binding.
 FT ACT_SITE 122 122 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 878 AA; 97588 MW; 6F7D181BADA617E3 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 878;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 DEAMLFSA 140
 DB 691 DEAMLFSA 698

RESULT 3
 ID RC_BPPH1 STANDARD; PRT; 144 AA.

AC P06153; P15239;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE IMMUNITY REPRESSOR PROTEIN.
 OS Bacteriophage phi-105.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10717;
 RN [1]

RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RX MEDLINE=88283656; PubMed=3135184;
 RA van Kaer L., Gansemans Y., van Montagu M., Dhaese P.;
 RT "Interaction of the Bacillus subtilis phage phi 105 repressor DNA: a
 RT genetic analysis.";
 RL EMBO J. 7:859-866(1988).

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85297750; PubMed=2993999;
 RA Dhaese P., Seurinck J., de Smet B., van Montagu M.;
 RT "Nucleotide sequence and mutational analysis of an immunity repressor
 RT gene from Bacillus subtilis temperate phage phi 105.";
 RL Nucleic Acids Res. 13:5441-5455(1985).

RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86056972; PubMed=3934047;
 RA Cully D.F., Garro A.J.;
 RT "Nucleotide sequence of the immunity region of Bacillus subtilis
 RT bacteriophage phi 105: identification of the repressor gene and its
 RT mRNA and protein products.";
 RL Gene 38:153-164(1985).

CC -!- SIMILARITY: BELONGS TO THE PBSX(XRE) FAMILY OF TRANSCRIPTIONAL
 CC REGULATORS.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X02799; CAA26567.1; ALT_INIT.
 DR EMBL; M11920; AAA88396.1; -

DR PIR; A24339; RBPPE5.
 DR PIR; S02459; S02459.
 DR InterPro: IPR001387; HTH_3.
 DR Pfam; PF01381; HTH_3; 1.
 DR SMART; SM00530; HTH_XRE; 1.
 KW Transcription regulation; Repressor; DNA-binding; Early protein.
 FT DNA_BIND 18 37
 H-T-H MOTIF (POTENTIAL).
 SQ SEQUENCE 144 AA; 16520 MW; EEED6EB3E4B34A1E CRC64;

Query Match 2.0%; Score 7; DB 1; Length 144;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 ETLIKEE 38
 DB 65 ETLIKEE 71

RESULT 4

ID SYB2_RHIME STANDARD; PRT; 151 AA.
 AC Q92301;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)

20-AUG-2001 (Rel. 40, Last annotation update)
 DE PROBABLE TRANSCRIPTIONAL REGULATOR SYR2.
 GN SYR2.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OG Plasmid pSymA (megaplasmid 1).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CXMI-105;
 RX MEDLINE=99413305; PubMed=10485295;
 RA Sharypova L.A., Yurgel S.N., Kellier M., Simarov B.V., Puehler A.,
 RA Becker A.;
 RT "The eff-482 locus of Sinorhizobium meliloti CXMI-105 that influences
 RT symbiotic effectiveness consists of three genes encoding an
 RT endoglycanase, a transcriptional regulator and an adenylate cyclase.";
 RL Mol. Gen. Genet. 261:1032-1044(1999).
 CC -!- FUNCTION: SEEMS TO AFFECT THE TRANSCRIPTION OF CVA3. MAY BE
 CC NEGATIVELY AUTOREGULATED.
 CC -!- SIMILARITY: BELONGS TO THE SYRB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ225896; CAB39102.1; -
 KW Nodulation; Transcription regulation; Repressor; Plasmid.
 SQ SEQUENCE 151 AA; 16694 MW; 08A85DC217812BFC CRC64;
 Query Match 2.0%; Score 7; DB 1; Length 151;
 Best Local Similarity 100.0%; Pred. No. 15;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 184 EGVKTLK 190
 |||||
 Db 73 EGVKTLK 79
 RESULT 5
 IF3_BUCAP STANDARD; PRT; 178 AA.
 AC P46243;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSLATION INITIATION FACTOR IF-3.
 GN INFC.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 CC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=98794;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95245268; PubMed=7766160;
 RA Kolibachuk D., Rouhakhsh D., Baumann P.;
 RT "Aromatic amino acid biosynthesis in Buchnera aphidicola (endosymbiont
 RT of aphids): cloning and sequencing of a DNA fragment containing
 RT aroH-thrs-infC-rpmI-rplT.";
 RL Curr. Microbiol. 30:313-316(1995).
 CC -!- FUNCTION: IF-3 BINDS TO THE 30S RIBOSOMAL SUBUNIT AND SHIFTS THE
 CC EQUILIBRIUM BETWEEN 70S RIBOSOMES AND THEIR 50S AND 30S SUBUNITS IN
 CC FAVOR OF THE FREE SUBUNITS, THUS ENHANCING THE AVAILABILITY OF 30S
 CC SUBUNITS ON WHICH PROTEIN SYNTHESIS INITIATION BEGINS.
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -!- SIMILARITY: BELONGS TO THE IF-3 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U11066; AAC43606.1; -
 DR HSSP; P02999; 2IFE.
 DR InterPro; IPR001288; IF3.
 DR Pfam; PF00707; IF3; 1.
 DR PRODOM; PD002880; IF3; 1.
 DR PROSITE; PS00938; IF3; 1.
 KW Initiation factor; Protein biosynthesis.
 FT SITE 106 106 IMPORTANT FOR 30S BINDING
 FT SITE 109 109 (BY SIMILARITY).
 FT SITE 109 109 IMPORTANT FOR 30S BINDING
 FT SITE 106 106 (BY SIMILARITY).
 SQ SEQUENCE 178 AA; 20601 MW; 9365C9D6512B1A3E CRC64;
 Query Match 2.0%; Score 7; DB 1; Length 178;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 172 LRELKKA 178
 |||||
 Db 39 LRELKKA 45
 RESULT 6
 YD37_METJA STANDARD; PRT; 231 AA.
 ID YD37_METJA
 AC Q58733;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HYPOTHETICAL PROTEIN MJ1337.
 GN MJ1337.
 OS Methanococcus jannaschii.
 CC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
 CC Methanococcus.
 OX NCBI_TaxID=2190;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
 RT jannaschii.";
 RL Science 273:1058-1073(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U67574; AAB99358.1; -
 DR TIGR; MJ1337; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 231 AA; 25674 MW; 203926DBC472FAE CRC64;

Query Match 2.0%; Score 7; DB 1; Length 231;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 309 GELAFILA 315
|||||||

Db 94 GELAFILA 100

RESULT 7

Y4MC_RHISN STANDARD; PRT; 237 AA.
AC P55562;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHELICAL PROTEIN Y4MC PRECURSOR.
GN Y4MC.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freilberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RL Nature 387:394-401(1997).
CC -!- SIMILARITY: NONE OBVIOUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000085; AAB91766.1; -
KW Hypothetical protein; Plasmid; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 237 HYPOTHETICAL PROTEIN Y4MC.
SQ SEQUENCE 237 AA; 26205 MW; F5AA67D9EB7BFBSB CRC64;

Query Match 2.0%; Score 7; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 83 PLSRVEA 89
|||||||

Db 41 PLSRVEA 47
[1]

RESULT 8

END8_MYCTU STANDARD; PRT; 255 AA.
AC P96902;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE ENDONUCLEASE VIII (EC 3.2.-.-).
GN NEI OR RV2297 OR MT3396 OR MTCY17.37.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=96295987; PubMed=9634230;

RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DNA N-GLYCOSYLASE WITH AN AP LASE ACTIVITY. REQUIRED
CC FOR THE REPAIR OF OXIDATIVE DNA DAMAGE (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE FPG FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z92771; CAB07061.1; -
DR EMBL; AE007148; AAK47739.1; -
DR TIGR; MT3396; -
DR TubercuList; RV3297; -
DR InterPro; IPR000191; Fapy_DNA_glyco.
DR InterPro; IPR000214; Fapy_DNAglyco_zn.
DR Pfam; PF01149; Fapy_DNA_glyco; 1.
DR ProDom; PD003680; Fapy_DNA_glyco; 1.
DR PROSITE; PS01242; FPG; 1.
KW DNA repair; Hydrolase; Glycosidase; Endonuclease; Zinc;
KW Complete proteome.
FT ZN_FING 230 253 POTENTIAL.
SQ SEQUENCE 255 AA; 28525 MW; 2C97B16CD7CD4002 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 255;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 253 GLLEVID 259
|||||||

Db 106 GLLEVID 112

RESULT 9

DAPA_XYLFA STANDARD; PRT; 302 AA.
AC Q9PER5;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DIHYDRODIPICOLINATE SYNTHASE (EC 4.2.1.52) (DHDPS).
GN DAPA OR XF0963.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.

Thu Apr 4 16:34:55 2002

STRAIN=9A5C;
 MEDLINE=20365717; PubMed=10910347;
 RX Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvares R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Bionnes M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
 RA Colauto N.B., Colombo C., Costa F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorfi H.,
 RA Faccinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kurane E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemes M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Martins A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 CC -!- CATALYTIC ACTIVITY: L-ASPARTATE 4-SEMIALDEHYDE + PYRUVATE =
 CC DIHYDRODIPICOLINATE + 2 H(2)O.
 CC -!- PATHWAY: BIOSYNTHESIS OF DIAMINOPIMELATE AND LYSINE FROM ASPARTATE
 CC SEMIALDEHYDE; FIRST STEP.
 CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE DHDPs FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF033934; AAF83773.1; ALT_INIT.
 CC InterPro: IPR002220; DHDPs.
 CC Pfam: PF00701; DHDPs; 1.
 CC PRINTS: PR00146; DHPICSWTHASE.
 CC ProDom: PD001859; DHDPs; 1.
 CC PROSITE: PS00665; DHDPs; 1.
 CC PROSITE: PS00666; DHDPs; 2; 1.
 CC Lyase; Diaminopimelate biosynthesis; Lysine biosynthesis;
 CC Complete proteome.
 CC ACT_SITE 162 BY SIMILARITY.
 CC SEQUENCE 302 AA; 31525 MW; 2298992B64D8EC70 CRC64;
 SQ

Query Match 2.0%; Score 7; DB 1; Length 302;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 79 LPLPLPLS 85
 Db 268 LPLPLPLS 274
 |||||
 |||||

RESULT 10
 ID MRAY_THEME STANDARD; PRT; 302 AA.
 AC Q9WY77;
 DT 30-MAY-2000 (Rel. 39, Created)

30-MAY-2000 (Rel. 39, Last sequence update)
 20-AUG-2001 (Rel. 40, Last annotation update)
 PHOSPHO-N-ACETYLURAMIDYL-PENTAPEPTIDE-TRANSFERASE (EC 2.7.8.13) (UDP-
 MURNAC-PENTAPEPTIDE PHOSPHOTRANSFERASE).
 DE MRAY OR TM0235.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogales; Thermotoga.
 OX NCBI_TaxID=2336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima";
 RL Nature 398:323-329(1999).
 CC -!- FUNCTION: FIRST STEP OF THE LIPID CYCLE REACTIONS IN THE
 CC BIOSYNTHESIS OF THE CELL WALL PEPTIDOGLYCAN (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: UDP-N-ACETYLMURAMYL-L-ALANYL-D-GLUTAMYL-L-
 CC LYSYL-D-ALANYL-D-ALANINE + UNDECAPRENYL PHOSPHATE = UMP +
 CC N-ACETYLMURAMYL-L-ALANYL-D-GLUTAMYL-L-ALANYL-D-ALANINE-
 CC DIPHOUPHOSPHATE.
 CC -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 4. MRAY
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF001707; AAD35326.1;
 CC TIGR: TM0235;
 CC InterPro: IPR000715; Glycos_transf_4.
 CC Pfam: PF00953; Glycos_transf_4; 1.
 CC PROSITE: PS01347; MRAY_1; 1.
 CC PROSITE: PS01348; MRAY_2; 1.
 CC Peptidoglycan synthesis; Cell division; Transferase; Transmembrane;
 CC Complete proteome.
 CC TRANSMEM 1 21 POTENTIAL.
 CC TRANSMEM 42 62 POTENTIAL.
 CC TRANSMEM 67 87 POTENTIAL.
 CC TRANSMEM 95 115 POTENTIAL.
 CC TRANSMEM 123 143 POTENTIAL.
 CC TRANSMEM 154 174 POTENTIAL.
 CC TRANSMEM 178 198 POTENTIAL.
 CC TRANSMEM 204 224 POTENTIAL.
 CC TRANSMEM 229 249 POTENTIAL.
 CC TRANSMEM 281 301 POTENTIAL.
 CC SEQUENCE 302 AA; 33814 MW; BB8FF74FEA9205CB CRC64;
 SQ

Query Match 2.0%; Score 7; DB 1; Length 302;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 28 IPVIETL 34
 Db 237 IPVIETL 243
 |||||
 |||||

RESULT 11
 ID CODV_BACSU STANDARD; PRT; 304 AA.
 DT 30-MAY-2000 (Rel. 39, Created)

```

AC P39776;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE INTEGRASE/RECOMBINASE CODV.
GN CODV.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=95302982; PubMed=7783641;
RA Slack F.J., Serror P., Joyce E., Sonenshein A.L.;
RT "A gene required for nutritional repression of the Bacillus subtilis
RT dipeptide permease operon.";
RL Mol. Microbiol. 15:689-702(1995).
RN (2)
RP SEQUENCE OF 1-85 FROM N.A.
RC STRAIN=168;
RA Foulger D., Errington J.;
RT "Cloning and sequencing 7.5Kbp of DNA from Bacillus subtilis upstream
RT of the codV gene.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE "PHAGE" INTEGRASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U13634; AAB03369.1; -
DR EMBL; AJ000975; CAA04424.1; -
DR EMBL; Z99112; CAB13487.1; -
DR HSP; P21891; LAOP.
DR Subtilist; BG10965; codV.
DR InterPro; IPR002104; Phage_integrase.
DR Pfam; PF00589; Phage_integrase; 1.
KW DNA recombination; DNA integration; Complete proteome.
FT ACT_SITE 281 TRANSIENT COVALENT LINKAGE TO DNA DURING
FT ACT_SITE 281 STRAND CLEAVAGE AND REJOINING (BY
FT SIMILARITY).
FT SEQUENCE 304 AA; 35342 MW; B0778BDA065A053 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 304;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 EELFEVS 104
Db 120 EELFEVS 126
|||||

RESULT 12
LPXK_RICRI STANDARD; PRT; 321 AA.
AC P58187;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE TETRAACYCLIDISACCHARIDE 4'-KINASE (EC 2.7.1.130) (LIPID A 4'-KINASE).
GN LPXK.
OS Rickettsia rickettsii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=783;
RN (1)
RP SEQUENCE FROM N.A.

P39776;
MEDLINE=21219194; PubMed=11319266;
Andersson J.O., Andersson S.G.E.;
"Pseudogenes, junk DNA and the dynamics of Rickettsia genomes.";
Mol. Biol. Evol. 18:829-839(2001).
-1- FUNCTION: TRANSFERS THE GAMMA-PHOSPHATE OF ATP TO THE 4'-POSITION
OF A TETRAACYCLIDISACCHARIDE 1-PHOSPHATE INTERMEDIATE (TERMED DS-1-
P) TO FORM TETRAACYCLIDISACCHARIDE 1,4'-BIS-PHOSPHATE (LIPID IVA)
(BY SIMILARITY).
-1- CATALYTIC ACTIVITY: ATP + 2,3-BIS(3-HYDROXYTETRADECANOYL)-D-
GLUCOSAMINYL-(BETA-D-1,6)-2,3-BIS(3-HYDROXYTETRADECANOYL)-D-
GLUCOSAMINYL BETA-PHOSPHATE = ADP + 2,3,2',3'-TETRAKIS(3-
HYDROXYTETRADECANOYL)-D-GLUCOSAMINYL-1,6-BETA-D-GLUCOSAMINE 1,4'-
BISPHOSPHATE.
-1- PATHWAY: LIPID A BIOSYNTHESIS; SIXTH STEP.
-1- SIMILARITY: BELONGS TO THE LPXK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ293329; CAC33716.1; -
KW Transferase; Kinase; Lipid A biosynthesis; Lipid synthesis;
KW ATP-binding. 54
FT NP_BIND 61 ATP (POTENTIAL).
FT SEQUENCE 321 AA; 36083 MW; 45E6CEB6CC50C0AE CRC64;

Query Match 2.0%; Score 7; DB 1; Length 321;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 41 GLNITGY 47
Db 247 GLNITGY 253
|||||

RESULT 13
LPXK_RICMO STANDARD; PRT; 325 AA.
AC P58186;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TETRAACYCLIDISACCHARIDE 4'-KINASE (EC 2.7.1.130) (LIPID A 4'-KINASE).
GN LPXK.
OS Rickettsia montana.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=33991;
RN (1)
RP SEQUENCE FROM N.A.
RC MEDLINE=21219194; PubMed=11319266;
Andersson J.O., Andersson S.G.E.;
"Pseudogenes, junk DNA and the dynamics of Rickettsia genomes.";
Mol. Biol. Evol. 18:829-839(2001).
-1- FUNCTION: TRANSFERS THE GAMMA-PHOSPHATE OF ATP TO THE 4'-POSITION
OF A TETRAACYCLIDISACCHARIDE 1-PHOSPHATE INTERMEDIATE (TERMED DS-1-
P) TO FORM TETRAACYCLIDISACCHARIDE 1,4'-BIS-PHOSPHATE (LIPID IVA)
(BY SIMILARITY).
-1- CATALYTIC ACTIVITY: ATP + 2,3-BIS(3-HYDROXYTETRADECANOYL)-D-
GLUCOSAMINYL-(BETA-D-1,6)-2,3-BIS(3-HYDROXYTETRADECANOYL)-D-
GLUCOSAMINYL BETA-PHOSPHATE = ADP + 2,3,2',3'-TETRAKIS(3-
HYDROXYTETRADECANOYL)-D-GLUCOSAMINYL-1,6-BETA-D-GLUCOSAMINE 1,4'-
BISPHOSPHATE.
-1- PATHWAY: LIPID A BIOSYNTHESIS; SIXTH STEP.
-1- SIMILARITY: BELONGS TO THE LPXK FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AJ293330; CAC33651.1; -
KW Transferase; Kinase; Lipid A biosynthesis; Lipid synthesis;
KW ATP-binding. 54 61 ATP (POTENTIAL)

FT NP_BIND 325 AA; 36481 MW; F59126EC871BD9FB CRC64;
SQ SEQUENCE 325 AA; 36481 MW; F59126EC871BD9FB CRC64;

Query Match 2.0%; Score 7; DB 1; Length 325;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 41 GLNITGY 47
Db 247 GLNITGY 253

RESULT 14
SYW_BACSU STANDARD; PRT; 330 AA.
AC P21656;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRIPTOPHANYL-TRNA SYNTHETASE (EC 6.1.1.2) (TRIPTOPHAN--TRNA LIGASE)
DE (TRPRS).
GN TRPS.

OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89211991; PubMed=3149612;
RA Chow K.C., Wong T.F.;
RT "Cloning and nucleotide sequence of the structural gene coding for
RT Bacillus subtilis tryptophanyl-trna synthetase.";
RL Gene 73:537-574(1988).

CC SEQUENCE OF 1-32.
CC MEDLINE=89174673; PubMed=2494170;
CC Xu Z.J., Love M.L., Ma L.Y.Y., Blum M., Bronskill P.M., Bernstein J.,
CC Grey A.A., Hofmann T., Camerman N., Wong J.T.F.;
CC "Tryptophanyl-trna synthetase from Bacillus subtilis.
CC Characterization and role of hydrophobicity in substrate
CC recognition.";
CC J. Biol. Chem. 264:4304-4311(1989).
CC -1- CATALYTIC ACTIVITY: ATP + L-TRIPTOPHAN + TRNA(TRP) = AMP +
CC PYROPHOSPHATE + L-TRIPTOPHANYL-TRNA(TRP).
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M24068; AAA22874.1; -
DR EMBL: Z99110; CAB12999.1; -
DR PIR: JT0481; YWES
DR Subtilisin; BG10799; trps.
DR InterPro: IPR002305; trna-synt_lb.
DR InterPro: IPR001412; trna-synt_I.

DR InterPro: IPR002306; trna-synt_trp.
DR Pfam: PF00579; trna-synt_lb; 1.
DR PRINTS: PR01039; TRNASYNTHTRP.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.
KW Aminoacyl-trna synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 11 19 "HIGH" REGION.
FT SITE 193 197 "KMSKS" REGION.
FT BINDING 196 196 ATP (BY SIMILARITY).
SQ SEQUENCE 330 AA; 37197 MW; 31927E2BC6320544 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 230 ILLYGTD 236
Db 134 ILLYGTD 140

RESULT 15
GBA3_USTHO STANDARD; PRT; 354 AA.
ID GBA3_USTHO
AC G14438;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE-BINDING PROTEIN ALPHA-3 SUBUNIT.
DE FILL.
GN FILL.

OS Ustilago hordei (Smut fungus).
OC Ascomycota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=120017;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=UHI100;
RX MEDLINE=98055158; PubMed=9393440;
RA Lichter A., Mills D.;
RT "Fill, a G-protein alpha-subunit that acts upstream of cAMP and is
RT essential for dimorphic switching in haploid cells of Ustilago
RT hordei.";
RL Mol. Gen. Genet. 256:426-435(1997).

CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE
CC INVOLVED AS MODULATORS OR TRANSDUCERS IN VARIOUS TRANSMEMBRANE
CC SIGNALING SYSTEMS. THIS SUBUNIT IS INVOLVED IN CAMP REGULATION AND
CC MORPHOGENESIS. IT IS ESSENTIAL FOR DIMORPHIC SWITCHING IN HAPLOID
CC CELLS.
CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).
CC THE ALPHA CHAIN CONTAINS THE GUANINE NUCLEOTIDE BINDING SITE.
CC -1- SIMILARITY: BELONGS TO THE G-ALPHA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U76672; AAC49880.1; -
DR InterPro: IPR001019; Gprotein_alpha.
DR Pfam: PF00503; G-alpha; 1.
DR PRINTS: PR00318; GPROTEINA.
DR SMART: SM00275; G-alpha; 1.
KW GTP-binding; Transducer; Multigene family.
FT NP_BIND 41 48 GTP (BY SIMILARITY).
FT NP_BIND 202 206 GTP (BY SIMILARITY).
FT NP_BIND 271 274 GTP (BY SIMILARITY).
SQ SEQUENCE 354 AA; 40405 MW; 6FB4D18305E76AC8 CRC64;

Query Match 2.0%; Score 7; DB 1; Length 354;

Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0;

Indels 0; Gaps 0;

QY 115 DP11PAI 121

|||||

Db 135 DP11PAI 141

Search completed: April 4, 2002, 14:50:10
Job time: 248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:45:32 ; Search time 28.85 Seconds
(without alignments)
1/54.254 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 346
Sequence: 1 LRALVFHGNLQYAEIPKSEP.....ERRLDAPRAIYNDWRCNGEP 346

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : SPTREMBL_17:*
- 1: sp.archaea:*
 - 2: sp.bacteria:*
 - 3: sp.fungi:*
 - 4: sp.human:*
 - 5: sp.invertebrate:*
 - 6: sp.mammal:*
 - 7: sp.mhc:*
 - 8: sp.organelle:*
 - 9: sp.phage:*
 - 10: sp.plant:*
 - 11: sp.rodent:*
 - 12: sp.virus:*
 - 13: sp.vertebrate:*
 - 14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	18	5.2	364	1 Q9HHB5	Q9hhb5 pyrococcus
2	12	3.5	364	1 Q58106	Q58106 pyrococcus
3	8	2.3	102	2 Q9K2M4	Q9k2m4 lactococcus
4	8	2.3	105	2 Q9LC53	Q9lc53 lactococcus
5	8	2.3	218	1 Q58562	Q58562 methanococ
6	8	2.3	243	2 P73217	P73217 synecocyst
7	8	2.3	262	5 O16952	O16952 caenorhabdi
8	8	2.3	272	1 Q59526	Q59526 pyrococcus
9	8	2.3	272	1 Q9V121	Q9v121 pyrococcus
10	8	2.3	328	5 O16953	O16953 caenorhabdi
11	8	2.3	466	2 Q30418	Q30418 lactococcus
12	8	2.3	466	2 Q9CG20	Q9cg20 lactococcus
13	8	2.3	492	10 Q80497	Q80497 arabidopsis
14	7	2.0	85	12 Q90992	Q90992 human immun
15	7	2.0	119	11 Q08929	Q08929 mus musculus
16	7	2.0	128	2 Q9CFK0	Q9cfk0 lactococcus
17	7	2.0	137	2 P96467	P96467 streptococ
18	7	2.0	147	9 Q80190	Q80190 bacterioph
19	7	2.0	160	10 Q9AVF7	Q9avf7 amaranthus

20	7	2.0	175	1 Q28131	Q28131 archaeoglob
21	7	2.0	195	2 Q9R5Y6	Q9r5y6 azotobacter
22	7	2.0	196	11 Q9D702	Q9d702 mus musculu
23	7	2.0	212	12 Q9WH18	Q9wh18 fish lympho
24	7	2.0	215	11 Q70315	Q70315 mus musculu
25	7	2.0	221	2 Q9X4K5	Q9x4k5 pseudomonas
26	7	2.0	241	12 P89296	P89296 yam mosaic
27	7	2.0	242	6 Q9N1Q5	Q9niq5 hylobates l
28	7	2.0	248	2 Q9KC51	Q9kc51 bacillus ha
29	7	2.0	252	10 Q9LGY4	Q9lgy4 oryza sativ
30	7	2.0	262	13 Q90713	Q90713 gallus gall
31	7	2.0	267	1 Q9V0E2	Q9v0e2 pyrococcus
32	7	2.0	271	2 Q9S4L4	Q9s4l4 streptococc
33	7	2.0	277	1 Q58047	Q58047 pyrococcus
34	7	2.0	277	2 Q9WXG7	Q9wxg7 alcaligenes
35	7	2.0	284	5 Q9ZHX1	Q9zhx1 brucella me
36	7	2.0	285	5 Q9U2H9	Q9u2h9 caenorhabdi
37	7	2.0	286	4 Q9WZ30	Q9wz30 homo sapien
38	7	2.0	291	4 Q9Y3A1	Q9y3a1 homo sapien
39	7	2.0	295	2 Q9CMB9	Q9cmb9 pasteurella
40	7	2.0	296	2 Q9A6J6	Q9a6j6 caulobacter
41	7	2.0	301	4 Q9NTX5	Q9ntx5 homo sapien
42	7	2.0	317	10 Q9AWS7	Q9aws7 oryza sativ
43	7	2.0	318	2 Q9PER5	Q9per5 xylella fas
44	7	2.0	325	4 Q9UFM6	Q9ufm6 homo sapien
45	7	2.0	325	4 Q9BTF9	Q9btf9 homo sapien

ALIGNMENTS

RESULT 1
Q9HHB5 PRELIMINARY; PRT; 364 AA.
AC Q9HHB5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE ALPHA-GALACTOSIDASE.
GN GALA.

OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3638;
RA Verhees C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF195244; AAC28455.1; 3E8E0AF5BDCCF2A5 CRC64;
SQ SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;

Query Match 5.2%; Score 18; DB 1; Length 364;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 RALVFHGNLQYAEIPKSE 19
DB 2 RALVFHGNLQYAEIPKSE 19
|||||

RESULT 2
Q58106 PRELIMINARY; PRT; 364 AA.
ID Q58106;
AC Q58106;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 41.8 KDA PROTEIN PH0368.
GN PH0368.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;

```

RN  SEQUENCE FROM N.A.
RP  STRAIN=OT3;
RC  MEDLINE=98344137; PubMed=9679194;
RA  Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA  Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA  Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA  Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA  Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA  Masuchi Y., Shizuya H., Kikuchi H.;
RT  "Complete sequence and gene organization of the genome of a hyper-
RT  thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL  DNA Res. 5:55-76(1998).
DR  EMBL; AP000002; BAA29442.1; -
KW  Hypothetical protein; Complete proteome.
SQ  SEQUENCE 364 AA; 41755 MW; 7B4B36AB4A975BAD CRC64;

Query Match      3.5%; Score 12; DB 1; Length 364;
Best Local Similarity 100.0%; Pred. No. 0.0014;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  6 FHGNLOYAEIPK 17
Db  ||||||
    6 FHGNLOYAEIPK 17

RESULT 3
ID Q9K2W4 PRELIMINARY; PRT; 102 AA.
AC Q9K2W4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE TRUNCATED GLUTAMATE DECARBOXYLASE.
GN GADB.
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ML, ATCC19257, 01-1, 924, F-16, AND HP;
RX MEDLINE=20250682; PubMed=10788408;
RA Nomura M., Kobayashi M., Ohmomo S., Okamoto T.;
RT "Inactivation of the glutamate decarboxylase gene in Lactococcus
RT lactis subsp. cremoris.";
RT Appl. Environ. Microbiol. 66:2235-2237(2000).
DR EMBL; AB033227; BAA95948.1; -
DR EMBL; AB033218; BAA95943.1; -
DR EMBL; AB033219; BAA95944.1; -
DR EMBL; AB033221; BAA95945.1; -
DR EMBL; AB033223; BAA95946.1; -
DR EMBL; AB033225; BAA95947.1; -
SQ SEQUENCE 102 AA; 11829 MW; 8854A05092CBE251 CRC64;

Query Match      2.3%; Score 8; DB 2; Length 102;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  295 DEGNARLN 302
Db  ||||||
    52 DEGNARLN 59

RESULT 4
ID Q9LC53 PRELIMINARY; PRT; 105 AA.
AC Q9LC53;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```

```

DE TRUNCATED GLUTAMATE DECARBOXYLASE (FRAGMENT).
GN GADB.
OS Lactococcus lactis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H-61;
RX MEDLINE=20250682; PubMed=10788408;
RA Nomura M., Kobayashi M., Ohmomo S., Okamoto T.;
RT "Inactivation of the glutamate decarboxylase gene in Lactococcus
RT lactis subsp. cremoris.";
RT Appl. Environ. Microbiol. 66:2235-2237(2000).
CC -!- CORFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC EMBL; AB033229; BAA95949.1; -
DR InterPro; IPR002129; Pyridoxal_dec.
DR Pfam; PF00282; Pyridoxal_dec. 1.
KW Decarboxylase; Lyase; Pyridoxal phosphate.
FT NON_TER 105 105
SQ SEQUENCE 105 AA; 12141 MW; E104D1BB1409D4EB CRC64;

Query Match      2.3%; Score 8; DB 2; Length 105;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  295 DEGNARLN 302
Db  ||||||
    52 DEGNARLN 59

RESULT 5
ID Q58562 PRELIMINARY; PRT; 218 AA.
AC Q58562;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ1162.
GN MJ1162.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kervatage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
DR EMBL; U67558; AAB99164.1; -
DR TIGR; MJ1162; -
DR InterPro; IPR000934; Ser_thr_phosphatse.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 218 AA; 24676 MW; 2F9D763416039EE4 CRC64;

Query Match      2.3%; Score 8; DB 1; Length 218;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  256 EVIDELNS 263

```



```

Db 70 EVIDELNS 77
|||||
RESULT 6
P73217 PRELIMINARY: PRT; 243 AA.
ID P73217
AC 01-FEB-1997 (TREMREL. 02, Created)
DT 01-FEB-1997 (TREMREL. 02, Last sequence update)
DR InterPro: IPR000276; GPCR_Rhodpsn.
DT 01-JUN-2000 (TREMREL. 14, Last annotation update)
DE HYPOTHETICAL 26.6 KDA PROTEIN.
GN SLL1570.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RA "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90904; BAA17244.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 26620 MW; 37DE109025773501 CRC64;

Query Match 2.3%; Score 8; DB 2; Length 243;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 218 PKKVASWI 225
Db 16 PKKVASWI 23
|||||

RESULT 7
O16952 PRELIMINARY: PRT; 262 AA.
ID O16952
AC O16952;
DT 01-JAN-1998 (TREMREL. 05, Created)
DT 01-JAN-1998 (TREMREL. 05, Last sequence update)
DR InterPro: IPR000276; GPCR_Rhodpsn.
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE COSMID R11G11.
GN R11G11.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Snelson J., Smith A., Sonnhammer E., Staden K., Waterston R.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wooldman P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [2]

SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Davidson S., Wohlmann P., Bauer C.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=BRISTOL N2;
RA Waterston R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF022976; AAC69084.1; -.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
SQ SEQUENCE 262 AA; 30310 MW; AD478C9F4F4344E2 CRC64;

Query Match 2.3%; Score 8; DB 5; Length 262;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 170 LLLRELK 177
Db 138 LLLRELK 145
|||||

RESULT 8
O59526 PRELIMINARY: PRT; 272 AA.
ID O59526
AC O59526;
DT 01-AUG-1998 (TREMREL. 07, Created)
DT 01-AUG-1998 (TREMREL. 07, Last sequence update)
DR InterPro: IPR000276; GPCR_Rhodpsn.
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE HYPOTHETICAL 31.7 KDA PROTEIN PH1894.
GN PH1894.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=OT3;
RA MEDLINE=98344137; PubMed=9679194;
RA Kwarabayasi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RA "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000007; BAA31016.1; -.
DR InterPro: IPR002309; trna_synt_2.
DR Pfam; PF01336; trna_antl; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 272 AA; 31654 MW; 3580BC7EEACCCDBB CRC64;

Query Match 2.3%; Score 8; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 254 LLEVIDEL 261
Db 173 LLEVIDEL 180
|||||

RESULT 9
O9V121 PRELIMINARY: PRT; 272 AA.
ID O9V121
AC O9V121;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DR InterPro: IPR000276; GPCR_Rhodpsn.
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE HYPOTHETICAL 31.8 KDA PROTEIN.

```

GN PAB2165.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORSAY;
 RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ248283; CAB49207.1; -
 DR InterPro: IPR002309; trna-synt_2.
 DR Pfam: PF01336; tRNA-anti; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 272 AA; 31827 MW; 515DA586372DD837 CRC64;

Query Match 2.3%; Score 8; DB 1; Length 272;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 254 LLEVIDEL 261
 Db 173 LLEVIDEL 180
 |||||

RESULT 10
 ID 016953 PRELIMINARY; PRT; 328 AA.
 AC 016953;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DE COSMID R11G11.
 GN R11G11.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Koopra A., Saunders D., Showkeen R.,
 RA Smalton N., Smith A., Sonnenhammer E., Staden K., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Davidson S., Wohlmann P., Bauer C.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Waterston R.;

Query Match 2.3%; Score 8; DB 5; Length 328;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 DEGNARLN 302
 Db 52 DEGNARLN 59
 |||||

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 170 LLLRELRLK 177
 Db 140 LLLRELRLK 147
 |||||

RESULT 11
 ID 030418 PRELIMINARY; PRT; 466 AA.
 AC 030418; OS0645;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DE 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE GLUTAMATE DECARBOXYLASE (EC 4.1.1.15).
 GN GADB.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN=MGI363;
 RX MEDLINE=98143417; PubMed=9484886;
 RA Sanders J.W., Leenhouts K., Burghoorn J., Brands J.R., Venema G.,
 RA Kok J.;
 RT "A chloride-inducible acid resistance mechanism in Lactococcus lactis and its regulation."
 RL Mol. Microbiol. 27:299-310(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=01-7;
 RA Nomura M., Fujita Y., Kobayashi M., Oomomo S.;
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CONVERTS INTERNALIZED GLUTAMATE TO GABA AND INCREASES THE INTERNAL PH. INVOLVED IN GLUTAMATE-DEPENDENT ACID RESISTANCE.
 CC -!- CATALYTIC ACTIVITY: L-GLUTAMATE - 4-AMINOBUTANOATE + CO(2).
 CC -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 CC -!- INDUCTION: EXPRESSION IS HIGHEST AT ONSET OF STATIONARY PHASE IN PRESENCE OF NaCl AND GLUTAMATE, AND AT LOW PH. CHLORIDE-DEPENDENT EXPRESSION IS ACTIVATED BY GADR.
 CC -!- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND TYRDC).

DR EMBL; AF005098; AAC46188.1; -
 DR EMBL; AB010789; BAA24585.1; -
 DR InterPro: IPR002129; Pyridoxal_dec.
 DR Pfam: PF00282; pyridoxal_dec; 1.
 DR PROSITE: PS00392; DDC_GAD_HDC_YDC; FALSE_NEG.
 KW Lyase; Decarboxylase; Pyridoxal phosphate; Multigene family.
 FT BINDING 277 277
 FT VARIANT 114 114 G -> E (IN STRAIN 01-7).
 FT VARIANT 116 117 IY -> PM (IN STRAIN 01-7).
 FT VARIANT 172 172 V -> I (IN STRAIN 01-7).
 FT VARIANT 185 185 R -> K (IN STRAIN 01-7).
 FT VARIANT 193 193 E -> D (IN STRAIN 01-7).
 FT VARIANT 353 353 Y -> F (IN STRAIN 01-7).
 FT VARIANT 371 371 A -> S (IN STRAIN 01-7).
 FT VARIANT 383 383 N -> D (IN STRAIN 01-7).
 FT VARIANT 444 444 D -> E (IN STRAIN 01-7).
 FT VARIANT 453 453 F -> Y (IN STRAIN 01-7).
 FT VARIANT 455 455 O -> E (IN STRAIN 01-7).
 SQ SEQUENCE 466 AA; 53849 MW; 28B94EB3FEEB168D CRC64;

Query Match 2.3%; Score 8; DB 2; Length 466;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 DEGNARLN 302
 Db 52 DEGNARLN 59
 |||||

Best Local Similarity 100.0%; Pred. No. 23; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 180 KLVEGKV 187
Db 292 KLVEGKV 299
|||||

RESULT 14

O90992 PRELIMINARY; PRT; 85 AA.
ID O90992
AC O90992
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE GP120 (FRAGMENT).
CN ENV.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-V11267;
RA Heydrickx L., Janssens W., Coppens S., Vereecken K., Willems B.,
RA Franssen K., Colebunders R., Vandenbruaene M., Van Der Groen G.;
RT "HIV-1 C2V3 env diversity among Belgian individuals."
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ228225; CAA12855.1; -
DR InterPro: IPR00077; GP120.
DR Pfam: PF00516; GP120; 1.
FT NON_TER 1 85
FT NON_TER 85
SQ SEQUENCE 85 AA; 9453 MW; 1D1C9745BBB8602C CRC64;

Query Match 2.0%; Score 7; DB 12; Length 85;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DRVKEEL 100
Db 75 DRVKEEL 81
|||||

RESULT 15

O08929 PRELIMINARY; PRT; 119 AA.
ID O08929
AC O08929
DT 01-JUL-1997 (TREMREL. 04, Created)
DT 01-JUL-1997 (TREMREL. 04, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE CEAL14 PROTEIN (FRAGMENT).
GN CEAL14.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE=LIVER;
RA Zimmermann W., Fischer B., Olsen A., Nedellec P., Thompson J.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y13561; CAA73896.1; -
DR InterPro: IPR003599; Ig.
DR SMART: SM00409; IG; 1.
KW Pregnancy.
FT NON_TER 1 1
FT NON_TER 119 119
SQ SEQUENCE 119 AA; 13635 MW; E4E6658159641235 CRC64;

Query Match 2.0%; Score 7; DB 11; Length 119;
Best Local Similarity 100.0%; Pred. No. 63;

RESULT 12

O9CG20 PRELIMINARY; PRT; 466 AA.
AC O9CG20;
DT 01-JUN-2001 (TREMREL. 17, Created)
DT 01-JUN-2001 (TREMREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE GLUTAMATE DECARBOXYLASE (EC 4.1.1.15).
GN GADB.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IL1403;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis.";
RL Genome Res. 0:0-0(2001).
CC -|- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -|- SIMILARITY: PYRIDOXAL-DEPENDENT DECARBOXYLASE FAMILY.
EMBL: AE006361; AAK05388.1; -
DR InterPro: IPR002129; Pyridoxal_dec.
DR Pfam: PF00282; Pyridoxal_dec; 1.
DR Complete proteome: Decarboxylase; Lyase; Pyridoxal phosphate.
KW SEQUENCE 466 AA; 53926 MW; BCDC732253E473C2 CRC64;

Query Match 2.3%; Score 8; DB 2; Length 466;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 295 DEGNARLN 302
Db 52 DEGNARLN 59
|||||

RESULT 13

O80497 PRELIMINARY; PRT; 492 AA.
AC O80497;
DT 01-NOV-1998 (TREMREL. 08, Created)
DT 01-NOV-1998 (TREMREL. 08, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE PUTATIVE GLYCOSYL HYDROLASE.
GN F16B22.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Rounsley S.D., Kaul S., Lin X., Ketchum K.A., Crosby M.L.,
RA Brandon R.C., Sykes S.M., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RT "Arabidopsis thaliana chromosome II BAC F16B22 genomic sequence."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC003672; AAC27459.1; -
DR HSP; P26221; 1TF4.
DR Mendel; 31158; Arath;1057;31158.
DR InterPro: IPR001701; Glyco_hydro_9.
DR Pfam: PF00759; Glyco_hydro_9; 1.
DR PROSITE: PS00592; GLYCOSYL_HYDROL_F9_1; 1.
KW Hydrolase.
SQ SEQUENCE 492 AA; 53591 MW; 6FE816C89048D620 CRC64;

Query Match 2.3%; Score 8; DB 10; Length 492;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 LKHSGRE 276

|||||

Db 70 LKHSGRE 76

Search completed: April 4, 2002, 14:49:45
Job time: 253 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:34:16 ; Search time 25.56 Seconds
(without alignments)
1002.714 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 1778
Sequence: 1 LRLVHFNGLQYAEIPKSEP.....ERRLDFAIRYNDWRNGEP 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq.1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1586	89.2	364	AAW34643	Thermotable alpha
2	118.5	6.7	649	AAW47504	Pyrococcus furiosu
3	100.5	5.7	653	AAW54870	Super heat resista
4	97.5	5.5	434	AAW82181	S. epidermidis ope
5	97.5	5.5	557	AAW82196	S. epidermidis ope
6	97.5	5.5	1396	AAW36871	Protein involved i
7	93.5	5.3	216	AAW29078	Arabidopsis thalia
8	93.5	5.3	218	AAW29077	Arabidopsis thalia
9	93.5	5.3	235	AAW29076	Arabidopsis thalia
10	93	5.2	377	AAW34045	Zea mays protein f
11	93	5.2	465	AAW88328	Cotton UDP glucose

12	93	5.2	473	21	AAW07943	A uridine diphosph
13	93	5.2	473	21	AAW07948	A uridine diphosph
14	93	5.2	473	21	AAW34044	Zea mays protein f
15	91	5.1	335	21	AAW52610	Helicobacter pylori
16	91	5.1	480	21	AAW16318	Pinus radiata UGP
17	91	5.1	510	22	AAW46303	H. pylori HPS068 p
18	91	5.1	594	22	AAW84334	Amino acid sequenc
19	88.5	5.0	403	21	AAW05758	Arabidopsis thalia
20	88.5	5.0	403	21	AAW49835	Arabidopsis thalia
21	88.5	5.0	406	21	AAW05757	Arabidopsis thalia
22	88.5	5.0	406	21	AAW49834	Arabidopsis thalia
23	88.5	5.0	422	21	AAW05756	Arabidopsis thalia
24	88.5	5.0	422	21	AAW49833	Arabidopsis thalia
25	88.5	5.0	492	21	AAW21554	Arabidopsis thalia
26	87.5	4.9	919	22	AAW83181	Corynebacterium th
27	87.5	4.9	1118	22	AAW8284	Rice magnesium che
28	87	4.9	329	22	AAW82183	S. epidermidis ope
29	87	4.9	459	21	AAW41936	Arabidopsis thalia
30	87	4.9	908	20	AAW33450	Chimeric Taq DNA p
31	86.5	4.9	513	21	AAW74583	Neisseria meningit
32	86.5	4.9	527	22	AAW99962	ERA binding domain
33	86.5	4.9	1218	20	AAW34680	Chlamydia pneumoni
34	86	4.8	832	19	AAW23962	Modified Taq polym
35	85.5	4.8	286	21	AAW35951	Human 3-hydroxyacy
36	85.5	4.8	301	21	AAW87599	Human fatty acid b
37	85.5	4.8	474	21	AAW07947	A uridine diphosph
38	85.5	4.8	748	21	AAW51060	Arabidopsis thalia
39	85	4.8	534	21	AAW16317	Eucalyptus grandis
40	84.5	4.8	512	21	AAW74582	Neisseria gonorrhoe
41	84.5	4.8	907	21	AAW36818	Arabidopsis thalia
42	84.5	4.8	912	21	AAW36817	Arabidopsis thalia
43	84	4.7	720	22	AAW62032	Recombinant P. fur
44	84	4.7	720	22	AAW62033	P. furiosus helica
45	84	4.7	832	19	AAW76205	T. aquaticus DNA p

ALIGNMENTS

RESULT 1
AAW34643
ID AAW34643 standard; Protein; 364 AA.

XX AC AAW34643;

XX DT 27-MAR-1998 (first entry)

XX DE Thermotable alpha-galactosidase AED1112RA-alpha-gal-18GC.

XX KW Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;

XX KW food processing; alpha glycoside hydrolysis; raffinose;

XX KW stachyose; verbascose; bean; flatulence; AED1112RA-alpha-gal-18GC.

XX OS Thermococcus alcaliphilus strain AED1112RA.

XX FH Key Location/Qualifiers

FT Misc-difference 329

FT /note= "encoded by CTT"

XX PN WO732974-A1.

XX PD 12-SEP-1997.

XX PF 05-FEB-1997; 97WO-US01452.

XX PR 08-MAR-1996; 96US-0613220.

XX PA (RECO-) RECOMBINANT BIOCATALYSIS INC.

XX PI Murphy D, Reid J, Rudolph MJ;

XX DR WPI: 1997-470541/43.

XX DR N-PSDB; AAT93753.

XX Nucleic acid encoding alpha-galactosidase from Thermococcus
 PT Alcaliphilus - used in food processing to hydrolyse
 PT alpha-glycosides, e.g. raffinose
 XX
 PS Claim 1; Fig 1; 32pp; English.
 XX
 CC This protein comprises ABDII12RA-alpha-gal-18GC, a claimed
 CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
 CC ABDII12RA, a bacterium that shows optimum growth at 85 deg C and
 CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
 CC AAT93753) encoding the alpha-galactosidase; (2) a vector containing
 CC the polynucleotide or homologous or complementary sequences; (2)
 CC host cells containing the vector; (3) a process for producing the
 CC alpha-galactosidase in transformed or transfected host cells; an
 CC enzyme showing at least 70% identity to alpha-galactosidase and
 CC comprising at least 30 amino acid residues of its sequence; and (4)
 CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
 CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
 CC in sugar beet processing (raffinose inhibits crystallisation of
 CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose
 CC and verbascone in beans and other gassy foods.
 XX
 SQ Sequence 364 AA;

Query Match 89.2%; Score 1586; DB 18; Length 364;
 Best Local Similarity 94.8%; Pred. No. 1.4e-150;
 Matches 345; Conservative 1; Mismatches 0; Indels 18; Gaps 18;
 QY 1 LRALVPHGNLQVAEIPKSE-PKVEIKAYIPVETLIKKE-PRGLNITGYTLKFLPKDIT- 57
 DB 1 LRALVPHGNLQVAEIPKSEIPKVEIKAYIPVETLIKKEIPFGLNITGYTLKFLPKDID 60
 QY 58 LVKGGIASDLIEIGTSYF-AILPLPLSRVQAQVQRDR-VKEELFEVSPKGFELPELA- 114
 DB 61 LVKGGIASDLIEIGTSYFHALPLPLSRVQAQVQRDRVKEELFEVSPKGFELPELAY 120
 QY 115 DFIIPAILKONGEYELFAD-EAMLFSAHNSAIKPIKPL-PHLIKRAQREKFRFYISYLL- 171
 DB 121 DPLIPAILKONGEYELFADGEAMLFSAHNSAIKPIKPLPHLYPHLIKAGREKFRFYISYLL 180
 QY 172 LRELKRAIKLVFEGKVTLK-VKDIKAVPWWAVNTAVML-IGRLPLMPKPKVASWIEDK- 228
 DB 181 LRELKRAIKLVFEGKVTLIKAVKDIEAVPWWAVNTAVMLIGRLPLMPKPKVASWIEDKD 240
 QY 229 NILLYGTDFIEGYRDIAG-RMSVEGLLEVIDELNSELC-PSELKHSGRELYLRTSSWA- 285
 DB 241 NILLYGTDFIEGYRDIAGRMSVEGLLEVIDELNSELCPLSELKHSGRELYLRTSSWAP 300
 QY 286 DKSRLRWREDEGNARLNML-YNMRGELAFLAENSARGW-PLPERRLDAPRAIYNDWRG- 342
 DB 301 DKSRLRWREDEGNARLNMLSYNMRGELAFLAENSARGWPELPERLDAPRAIYNDWRGE 360
 QY 343 NGEP 346
 DB 361 NGEP 364

RESULT 2
 AAR47504
 ID AAR47504 standard; protein; 649 AA.
 XX
 AC AAR47504;
 XX
 DT 07-JUL-1994 (first entry)
 XX
 DE Pyrococcus furiosus alpha amylase.
 XX
 KW Pyrococcus furiosus; alpha amylase; liquefaction; polymers;
 XX glucopolymers; thermostable.
 OS Pyrococcus furiosus.

XX EP577257-A.
 XX 05-JAN-1994.
 PD
 XX 17-MAY-1993; 93EP-0303801.
 XX
 PF 09-JUN-1992; 92US-0893928.
 PR
 XX (UWJO) UNIV JOHNS HOPKINS.
 PA
 XX Anfinsen CB, Laderman K;
 PI WPI; 1994-009532/02.
 DR
 XX Purified Pyrococcus furiosus alpha-amylase - used for the
 PT industrial liquefaction of gluco-polymers at high temps.
 XX
 PS Claim 2; Figure 9; 4lpp; English.
 XX
 CC The purified Pyrococcus furiosus alpha amylase can act on substrates
 CC with a low degree of polymerisation. e.g. glucose polymers as short
 CC as maltotriose. The enzyme can be used for efficient industrial
 CC liquefaction of glucopolymers at high temperatures.
 XX
 SQ Sequence 649 AA;

Query Match 6.7%; Score 118.5; DB 15; Length 649;
 Best Local Similarity 20.2%; Pred. No. 0.0052;
 Matches 79; Conservative 64; Mismatches 117; Indels 131; Gaps 21;
 QY 24 EKAYIPVETLIKKEP---EGLNITGYTLKFL---PKDILYKGGIASDLIEIGTS-Y 75
 DB 29 ekaywplfletl-eeypmkaivaitsgpliewlqdirpeyldlrsvkrqgveivvagfy 87
 QY 76 TAILPLPLSRVQAQVQRDRVKEELFE---VSPKGFELPE-LADPIIPAILKONGEYELF 131
 DB 88 epviasip---kedriegirlmkewaksigfdargvwltervwqpeivktlkesgidyvi 144
 QY 132 ADEAMLFSAHNS-----AIKPIKPLPHLIKRAQREKFRFYISYLLREL 175
 DB 145 vdyhfmssaelkeelywpyytedggevtafvpl-----dekrlryl---ipfrpv 191
 QY 176 RKAIKLVF-----EGKVTLKVKDIEAVPWWAVNTAVMLIGRLPLMPKPKVASWIEDKN 229
 DB 192 dkveylhslidgdeskvavfhddgekfgiwp-----tyewvyekg 233
 QY 230 ILLYGTDFIEGYRDIAGRMSVEGLLEVIDELNSELCPSSEL-KHSGREL-----YLRT 281
 DB 234 wl-----reffdrissd-----eklnlmlyteylekykprglvipliasyfm 276
 QY 282 SSWA--DKSLRWREDEGNARLNMLYN-----MRGEL-----AF 313
 DB 277 sewslpakqarlifvefnelkvkgeikyrvfvrgigwknffkykypesnmhkrmlmvsK 336
 QY 314 LAENSARGWPLPERLDAPRAIYND--WRG 342
 DB 337 lvrrnn-----pearckyllraqcndaywhg 360

RESULT 3
 AAW54870
 ID AAW54870 standard; Protein; 653 AA.
 XX
 AC AAW54870;
 XX
 DT 01-SEP-1998 (first entry)
 XX
 DE Super heat resistant 4-alpha-glucanotransferase.
 XX
 KW Super heat-resistant 4-alpha-glucanotransferase; heat-treatment;
 KW alpha-1,4-glucan; alpha-1,4-glucoside bond.

```

XX OS Pyrococcus sp.
XX JP10150986-A.
XX 09-JUN-1998.
XX 21-NOV-1996; 96JP-03111117.
XX 21-NOV-1996; 96JP-03111117.
XX (BEAB-) BE ABLE KK.
XX (NAGA-) NAGASE SEIKAGAKU KOGYO KK.
XX WPI; 1998-379989/33.
XX N-PSDB; AAV27026.
XX New 4-alpha-glucanotransferase which has been heat-treated - used at
XX high temperatures to transfer at least one glucose unit
XX Claim 4; Page 10-13; 18pp; Japanese.
XX The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of
XX 6.0-8.0, with an optimum temperature at pH 7.5 of 100 deg. C. It has at
XX least 90% activity after heat-treatment at 100 deg. C for 30 minutes at
XX pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glucan
XX to alpha-1,4-glucan by an alpha-1,4-glucoside bond.
XX Sequence 653 AA;

Query Match 5.7%; Score 100.5; DB 19; Length 653;
Best Local Similarity 19.3%; Pred. No. 0.33;
Matches 74; Conservative 67; Mismatches 127; Indels 115; Gaps 19;

QY 24 EKAYIPVIEFLKEEP---FGLNITGYTLKFL---PKDIIIVKGGIASDLIEIGTS-Y 75
bb 28 ersyrpfmetl-eeypnmkvavhysgplleiwirdnkpehlldlrsivkrgqleivvagfy 86
QY 76 TAILPLPLSRVAQVORDRKE--ELFEVSPKGFNLPE-LADPIITAILKONGEYELFA 132
Db 87 epwlasip--kedrivqieklikefarnlgyeargvwlttervwqpelvkslraagidyv 144
QY 133 DEAMLFSAHLNS-----AIKPIKPLPLIKAKQREKRRFRYSVLLLELR 176
Db 145 ddyhfsagskldelfwpytyedggevityvfp1-----deklyl--ipfrpd 191
QY 177 KAIKLVP-----EGKVLKVKDIEAVPVWVAVNTAVMLIGRLPLMNPKKVASWIEDK-- 228
Db 192 ktleylslldgdgskavvfhddgekfgvpg-----tyewvyekgw 233
QY 229 -----NLLYGTDTIE-----FIGYRDIAG--RMSVEGL-----LEVID 259
Db 234 lreffdrvsderinlmlyseylqtrfprglvlypiasyfemsewsiparqaklfefve 293
QY 260 ELNSELCPSELKHSRELYLRTSSWADKSLRIWREDEGNARLMNLYNMGRGELAFLENSD 319
Db 294 elk-----kenkforyrvfrrvgglwknffkypesnymhkrmlmvska-----vrnnp 342
QY 320 ARGWPLPERLDAPRAIYNDRG 342
Db 343 arefillraqnday-----whg 359

RESULT 4
AAG82181
ID AAG82181 standard; Protein; 434 AA.
XX AAG82181;
XX 03-SEP-2001 (first entry)
XX S. epidermidis open reading frame protein sequence SEQ ID NO:1456.

```

```

XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
XX vaccination; endocarditis.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US30782.
XX 09-NOV-1999; 99US-0164258.
XX (GLAX ) GLAXO GROUP LTD.
XX PA Kimmerly WJ;
XX PI WPI; 2001-316495/33.
XX DR N-PSDB; AAH53031.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 18; Page 411; 2188pp; English.
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
XX (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis
XX (I) and (II) can have antibacterial activity and therefore can be used
XX in vaccination. The nucleic acids (I) may be used to produce the
XX S. epidermidis polypeptides (II) via the production of vectors
XX containing them which are used to produce hosts cells which express the
XX polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
XX used to vaccinate subjects and to raise antibodies against the bacteria.
XX The polypeptides may also be used to assay for other inhibitors of their
XX activity and therefore identify compounds that may be used for the
XX treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
XX AAH55090 represent specifically claimed S. epidermidis genomic DNA
XX polynucleotide sequences from the present invention. AAH55091 to
XX AAH55098 represent oligonucleotide sequences and primers which are used
XX in the exemplification of the present invention.
XX N.B. The present invention specifically claims all the polynucleotide
XX sequences given in the sequence listing of the present specification,
XX however the sequence listing only goes up to SEQ ID NO:4454 so even
XX though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 434 AA;

Query Match 5.5%; Score 97.5; DB 22; Length 434;
Best Local Similarity 22.4%; Pred. No. 0.37;
Matches 81; Conservative 53; Mismatches 142; Indels 85; Gaps 16;

QY 13 AEPKSE-PKVIEKAYIPVIEFLKEEPFGLNITGYTLKFLPKD---IILVKGGIASDLI 68
Db 6 agvkserkkmkskqetlnkeplvkr--glkggyvveyrtdrtltievkmkaenga 63
QY 69 EIITGSYTAILPLPLSRVAQVORDRKEELFEVSPK-----GFWLPELADPIIPAIL 122
Db 64 eiinytksehftysdnkknvgievldmidgetyiaakkvinasgppwdev----- 114
QY 123 KONGEYELFADEAMLFSAHLNSAIKPIK-PLPHLIIKAKQREKFRYSVLLLEPR----- 174
Db 115 ---rsdyarnnkqrlrtkgvvhvldqskiplgqavfydtekdgrrmi-faipregkaygt 171
QY 175 -----LRKATIKLVFEGKVTLLKVKDIEAVPVWVAVNTAVNTAVNTAVNTAVNTAV 213
Db 172 tdtfydnekatplttqetdrdylnainymfp---tnvkdiediestwagirlpilekqkd 228
QY 214 PLMNPKKVASWIEDKNLLYGTDTIEFIGYRDIAGRMSVEGLLEVIDELNSELCPSELKHS 273
Db 229 pseisrkdew-egesglitiaggkltgyrhma-----leivdl-----akrl 271

```


PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 21-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145313.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 16-AUG-1999; 99US-0148684.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 30-AUG-1999; 99US-0151080.
 PR 31-AUG-1999; 99US-0151303.
 PR 01-SEP-1999; 99US-0151438.
 PR 07-SEP-1999; 99US-0151930.
 PR 10-SEP-1999; 99US-0152363.
 PR 13-SEP-1999; 99US-0153070.
 PR 15-SEP-1999; 99US-0153758.
 PR 16-SEP-1999; 99US-0154018.
 PR 20-SEP-1999; 99US-0154039.
 PR 22-SEP-1999; 99US-0154779.
 PR 23-SEP-1999; 99US-0155139.
 PR 24-SEP-1999; 99US-0155486.
 PR 28-SEP-1999; 99US-0155659.
 PR 29-SEP-1999; 99US-0156458.
 PR 04-OCT-1999; 99US-0156596.
 PR 05-OCT-1999; 99US-0157117.
 PR 06-OCT-1999; 99US-0157753.
 PR 07-OCT-1999; 99US-0157865.
 PR 08-OCT-1999; 99US-0158029.
 PR 12-OCT-1999; 99US-0158232.
 PR 13-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 18-OCT-1999; 99US-0159638.
 PR 21-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 28-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 5.3%; Score 93.5; DB 21; Length 216;
 Best Local Similarity 21.7%; Pred. No. 0.33; Indels 41; Gaps 7;
 Matches 45; Conservative 34; Mismatches 87;

QY 157 KAQREKFRFYISYLLRLRLRAIKLVFEGKVTLKVKDIEAVPVVAVNTAVMLIGRLPLM 216
 Db ksnqkfrfs-----eeqiksllelifesetrlpr-----kkvqvarelgl 47
 QY 217 NPKKVASWIEDKNILLYGTDIE-----FIGYROIAGRMV-----EGLLLEVIDELNSEL 265
 Db qprqvaiwfnkrarwtkkiekeyntlrannylniasqfelmkkeksglsvselqrlneem 107
 QY 266 -CPSELKH-----SGRELYLRTSSWADKSLRIWREDEGNARLN---MLYNMRGELAFLA 315
 Db 108 qprkeekhheccgdqglalssteshngksepegrldqgsvlncdgdynnnikteyfgfe 167
 QY 316 ENSDARGWPLPERRLDAFRAIYNDWRG 342
 Db 168 eetdheilmivekaddscsltssennwgg 194

RESULT 8

AAG29077
 ID AAG29077 standard; Protein; 218 AA.

XX AC AAG29077;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 34532.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 XX termination sequence.

XX OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0132956.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134376.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

XX 06-SEP-2000. 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
XX 06-JUL-1999; 99US-0142803.
PF 25-FEB-2000; 2000EP-0301439. PR 08-JUL-1999; 99US-0142920.
XX 25-FEB-1999; 99US-0121825. PR 12-JUL-1999; 99US-0142977.
PR 05-MAR-1999; 99US-0123180. PR 13-JUL-1999; 99US-0143542.
PR 09-MAR-1999; 99US-0123548. PR 14-JUL-1999; 99US-0143624.
PR 23-MAR-1999; 99US-0125788. PR 15-JUL-1999; 99US-0144005.
PR 25-MAR-1999; 99US-0126264. PR 16-JUL-1999; 99US-0144085.
PR 29-MAR-1999; 99US-0126785. PR 19-JUL-1999; 99US-0144086.
PR 01-APR-1999; 99US-0127462. PR 19-JUL-1999; 99US-0144325.
PR 06-APR-1999; 99US-0128234. PR 19-JUL-1999; 99US-0144331.
PR 08-APR-1999; 99US-0128714. PR 19-JUL-1999; 99US-0144332.
PR 16-APR-1999; 99US-0129845. PR 19-JUL-1999; 99US-0144333.
PR 19-APR-1999; 99US-0130077. PR 19-JUL-1999; 99US-0144334.
PR 21-APR-1999; 99US-0130449. PR 20-JUL-1999; 99US-0144335.
PR 23-APR-1999; 99US-0130510. PR 20-JUL-1999; 99US-0144352.
PR 23-APR-1999; 99US-0130891. PR 20-JUL-1999; 99US-0144632.
PR 28-APR-1999; 99US-0131449. PR 20-JUL-1999; 99US-0144884.
PR 30-APR-1999; 99US-0132048. PR 21-JUL-1999; 99US-0144814.
PR 30-APR-1999; 99US-0132407. PR 21-JUL-1999; 99US-0145086.
PR 04-MAY-1999; 99US-0132484. PR 21-JUL-1999; 99US-0145088.
PR 05-MAY-1999; 99US-0132485. PR 22-JUL-1999; 99US-0145085.
PR 06-MAY-1999; 99US-0132486. PR 22-JUL-1999; 99US-0145087.
PR 06-MAY-1999; 99US-0132487. PR 22-JUL-1999; 99US-0145089.
PR 07-MAY-1999; 99US-0132863. PR 22-JUL-1999; 99US-0145192.
PR 11-MAY-1999; 99US-0134256. PR 23-JUL-1999; 99US-0145145.
PR 14-MAY-1999; 99US-0134258. PR 23-JUL-1999; 99US-0145218.
PR 14-MAY-1999; 99US-0134219. PR 23-JUL-1999; 99US-0145224.
PR 14-MAY-1999; 99US-0134221. PR 26-JUL-1999; 99US-0145276.
PR 14-MAY-1999; 99US-0134370. PR 27-JUL-1999; 99US-0145913.
PR 18-MAY-1999; 99US-0134768. PR 27-JUL-1999; 99US-0145918.
PR 19-MAY-1999; 99US-0134941. PR 27-JUL-1999; 99US-0145919.
PR 20-MAY-1999; 99US-0135124. PR 28-JUL-1999; 99US-0145951.
PR 21-MAY-1999; 99US-0135353. PR 02-AUG-1999; 99US-0146386.
PR 24-MAY-1999; 99US-0135629. PR 02-AUG-1999; 99US-0146388.
PR 25-MAY-1999; 99US-0136021. PR 02-AUG-1999; 99US-0146389.
PR 27-MAY-1999; 99US-0136392. PR 03-AUG-1999; 99US-0147038.
PR 28-MAY-1999; 99US-0136782. PR 04-AUG-1999; 99US-0147204.
PR 01-JUN-1999; 99US-0137222. PR 04-AUG-1999; 99US-0147302.
PR 03-JUN-1999; 99US-0137528. PR 05-AUG-1999; 99US-0147192.
PR 04-JUN-1999; 99US-0137502. PR 05-AUG-1999; 99US-0147260.
PR 07-JUN-1999; 99US-0137724. PR 06-AUG-1999; 99US-0147303.
PR 08-JUN-1999; 99US-0138094. PR 06-AUG-1999; 99US-0147416.
PR 10-JUN-1999; 99US-0138540. PR 09-AUG-1999; 99US-0147493.
PR 10-JUN-1999; 99US-0138847. PR 09-AUG-1999; 99US-0147935.
PR 14-JUN-1999; 99US-0139119. PR 10-AUG-1999; 99US-0148171.
PR 16-JUN-1999; 99US-0139452. PR 11-AUG-1999; 99US-0148319.
PR 16-JUN-1999; 99US-0139453. PR 12-AUG-1999; 99US-0148341.
PR 17-JUN-1999; 99US-0139492. PR 13-AUG-1999; 99US-0148565.
PR 18-JUN-1999; 99US-0139454. PR 16-AUG-1999; 99US-0148684.
PR 18-JUN-1999; 99US-0139455. PR 17-AUG-1999; 99US-0149175.
PR 18-JUN-1999; 99US-0139456. PR 18-AUG-1999; 99US-0149426.
PR 18-JUN-1999; 99US-0139457. PR 20-AUG-1999; 99US-0149722.
PR 18-JUN-1999; 99US-0139458. PR 20-AUG-1999; 99US-0149723.
PR 18-JUN-1999; 99US-0139459. PR 20-AUG-1999; 99US-0149929.
PR 18-JUN-1999; 99US-0139460. PR 23-AUG-1999; 99US-0149930.
PR 18-JUN-1999; 99US-0139461. PR 23-AUG-1999; 99US-0150566.
PR 18-JUN-1999; 99US-0139462. PR 25-AUG-1999; 99US-0150884.
PR 18-JUN-1999; 99US-0139463. PR 26-AUG-1999; 99US-0151065.
PR 18-JUN-1999; 99US-0139750. PR 27-AUG-1999; 99US-0151065.
PR 18-JUN-1999; 99US-0139763. PR 27-AUG-1999; 99US-0151066.
PR 21-JUN-1999; 99US-0139817. PR 27-AUG-1999; 99US-0151080.
PR 22-JUN-1999; 99US-0139899. PR 30-AUG-1999; 99US-0151303.
PR 23-JUN-1999; 99US-0140353. PR 31-AUG-1999; 99US-0151438.
PR 23-JUN-1999; 99US-0140354. PR 01-SEP-1999; 99US-0151930.
PR 24-JUN-1999; 99US-0140695. PR 07-SEP-1999; 99US-0152363.
PR 28-JUN-1999; 99US-0140823. PR 10-SEP-1999; 99US-0153070.
PR 29-JUN-1999; 99US-0140991. PR 13-SEP-1999; 99US-0153758.
PR 30-JUN-1999; 99US-0141287. PR 15-SEP-1999; 99US-0154018.
PR 01-JUL-1999; 99US-0141842. PR 16-SEP-1999; 99US-0154039.
PR 01-JUL-1999; 99US-0142154. PR 20-SEP-1999; 99US-0154779.


```

PR 10-FEB-1999; 99US-0119588.
PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Famodu OO, Pearlstein RW;
XX
XX WPI: 2000-533019/48.
XX N-PSDB; AAA59672.
XX
XX New polynucleotides encoding uridine diphosphate (UDP) glucose
XX phosphorylase in plants and seeds, useful for creating transgenic
XX plants which express UDP glucose phosphorylase at higher or lower
XX levels than normal -
XX
XX Claim 1; Page 37-38; 46pp; English.
XX
XX The present sequence represents a plant uridine diphosphate (UDP) glucose
XX phosphorylase. The polynucleotides may be used to create transgenic
XX plants which express the UDP glucose phosphorylase at higher or lower
XX levels than normal or in cell types or developmental stages in which
XX they are not normally found. This would alter the level of cell wall
XX and starch biosynthesis in those cells. The polynucleotides may also
XX be used as probes for genetically and physically mapping the genes
XX that they are a part of, and as markers for traits linked to those
XX genes.
XX
XX SQ Sequence 473 AA;

Query Match 5.2%; Score 93; DB 21; Length 473;
Best Local Similarity 25.3%; Pred. No. 1.2;
Matches 59; Conservative 24; Mismatches 80; Indels 70; Gaps 12;

QY 53 PKDILVKGGA-SDL-----IEIIGTSYTAILPLPL-----SRVEA 89
DB 102 pksvievrngftflidivigieslnkkygcnvplllmnsfnthdtdqkivekysnsniei 161
QY 90 -----QVQRDRVKEELFEVSP-----KGFWLPELADPIIPA-----ILKDNQYEVLF 131
DB 162 htfnsgqyprivtedflplpsksgkdgwypgghgdfvpslnnsgkldillagkeyvf 221
QY 132 ADEAMLFSAHLNSAIKPKPLPHLIKAQREKFRFYISYLLRLRLKAIKLVFEGKVTL-- 189
DB 222 van-----sdnl-gaivdikihlhltqnqeycme-vtpkltadvkggtllisyegrqvle 275
QY 190 -----KVKDIEAVPV-----WVAVNTAVMLIGRLPLM-----NPKKV 221
DB 276 laqvpdehvnfksiekfifntnnlwnlkaivrlveaalkmeilpnkev 328

RESULT 13
AAB07948
ID AAB07948 standard; Protein; 473 AA.
XX
XX AC AAB07948;
XX
XX DT 14-NOV-2000 (first entry)
XX
XX DE A uridine diphosphate (UDP) glucose phosphorylase.
XX
XX KW Uridine diphosphate glucose phosphorylase; transgenic plant;
XX KW UDP glucose phosphorylase; cell wall synthesis; starch biosynthesis.
XX
XX OS Triticum aestivum.
XX
XX PN W0200047748-A2.
XX
XX PD 17-AUG-2000.
XX
XX PF 09-FEB-2000; 2000WC-US03513.
XX
XX PR 10-FEB-1999; 99US-0119588.
XX

```

```

PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
XX PI Famodu OO, Pearlstein RW;
XX
XX WPI: 2000-533019/48.
XX N-PSDB; AAA59677.
XX
XX New polynucleotides encoding uridine diphosphate (UDP) glucose
XX phosphorylase in plants and seeds, useful for creating transgenic
XX plants which express UDP glucose phosphorylase at higher or lower
XX levels than normal -
XX
XX Claim 1; Page 45-46; 46pp; English.
XX
XX The present sequence represents a plant uridine diphosphate (UDP) glucose
XX phosphorylase. The polynucleotides may be used to create transgenic
XX plants which express the UDP glucose phosphorylase at higher or lower
XX levels than normal or in cell types or developmental stages in which
XX they are not normally found. This would alter the level of cell wall
XX and starch biosynthesis in those cells. The polynucleotides may also
XX be used as probes for genetically and physically mapping the genes
XX that they are a part of, and as markers for traits linked to those
XX genes.
XX
XX SQ Sequence 473 AA;

Query Match 5.2%; Score 93; DB 21; Length 473;
Best Local Similarity 24.9%; Pred. No. 1.2;
Matches 58; Conservative 25; Mismatches 80; Indels 70; Gaps 12;

QY 53 PKDILVKGGA-SDL-----IEIIGTSYTAILPLPL-----SRVEA 89
DB 102 pksvievrngftflidivigieslnkkygcnvplllmnsfnthdtdqkivekysnsniei 161
QY 90 -----QVQRDRVKEELFEVSP-----KGFWLPELADPIIPA-----LKDNGYEVLF 131
DB 162 htfnsgqyprivtedflplpsksgkdgwypgghgdfvpslnnsgkldtllsqgkeyvf 221
QY 132 ADEAMLFSAHLNSAIKPKPLPHLIKAQREKFRFYISYLLRLRLKAIKLVFEGKVTL-- 189
DB 222 van-----sdnl-gaivdikihlhltqnqeycme-vtpkltadvkggtllisyegrqvle 275
QY 190 -----KVKDIEAVPV-----WVAVNTAVMLIGRLPLM-----NPKKV 221
DB 276 laqvpdehvnfksiekfifntnnlwnlkaivrlveaalkmeilpnkev 328

RESULT 14
AAG34044
ID AAG34044 standard; Protein; 473 AA.
XX
XX AC AAG34044;
XX
XX DT 18-OCT-2000 (first entry)
XX
XX DE Zea mays protein fragment SEQ ID NO: 41366.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence; corn.
XX
XX OS Zea mays subsp. mays.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123548.
XX

```


PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 30-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 02-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.

QY 1 LRALVFHGNLQVAEIPKSPKVIKAYIPVETLIKEEFGNITCYTLKFLPKDIIILVK 60
 Db 1 LRALVFHGNLQVAEIPKSPKVIKAYIPVETLIKEEFGNITCYTLKFLPKDIIILVK 60
 QY 61 GGTASDIEIIGTSYTAILPLPLSRVEAQVQDRVKEELFEVSPKGFWLPDELADPIIPA 120
 Db 61 GGTASDIEIIGTSYTAILPLPLSRVEAQVQDRVKEELFEVSPKGFWLPDELADPIIPA 120
 QY 121 ILKDNQYEVLFDAEAMLSAHLNSAIKPKPLPHLIKAQREKRFYISYLLRLRLKAIK 180
 Db 121 ILKDNQYEVLFDAEAMLSAHLNSAIKPKPLPHLIKAQREKRFYISYLLRLRLKAIK 180
 QY 181 LVFEGKVTILKVKDIEAVPVAVNTAVMLIGRLPLMNPCKKVASWIEDKNILLYGTDIEFI 240
 Db 181 LVFEGKVTILKVKDIEAVPVAVNTAVMLIGRLPLMNPCKKVASWIEDKNILLYGTDIEFI 240
 QY 241 GYRDIAGRMSVGLLEVIDELNSLCPSELKHSRGRELYLRTSSWADKSLRIWREDEGNAR 300
 Db 241 GYRDIAGRMSVGLLEVIDELNSLCPSELKHSRGRELYLRTSSWADKSLRIWREDEGNAR 300
 QY 301 LNNLNMKGELAFLENSDARGWPLPERRLDAFRAIYNDRGNGEP 346
 Db 301 LNNLNMKGELAFLENSDARGWPLPERRLDAFRAIYNDRGNGEP 346

RESULT 2
 US-07-894-212A-8
 ; Sequence 8, Application US/07894212A
 ; Patent No. 5366883
 ; GENERAL INFORMATION:
 ; APPLICANT: ASADA, KIYOZO
 ; APPLICANT: UEMORI, TAKASHI
 ; APPLICANT: MUKAI, HIROYUKI
 ; APPLICANT: KATO, IKUNOSHIN
 ; APPLICANT: LADERMAN, KENNETH
 ; APPLICANT: ANFINSEN, CHRISTIAN
 ; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CUSHMAN DARB & CUSHMAN
 ; STREET: 1100 NEW YORK AVENUE, N.W.
 ; CITY: WASHINGTON, D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/894,212A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KOKULIS, PAUL N.
 ; REGISTRATION NUMBER: 16773
 ; REFERENCE/DOCKET NUMBER: 95469/C-1195
 ; TELEPHONE: (202) 861-3000
 ; TELEFAX: (202) 822-0944
 ; TELEFAX: 6714627 CUSH
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 647 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-07-894-212A-8

Query Match 6.6%; Score 116.5; DB 1; Length 647;
 Best Local Similarity 20.2%; Pred. No. 0.0018;

* . . *

Matches 79; Conservative 64; Mismatches 117; Indels 131; Gaps 21;
 QY 24 EKAYIPVETLIKEE--FGLNITGYTLKFL-----PKDIIILKGGTASDIEIIGTS-Y 75
 Db 27 EKCVWPLETL--EYPMKVAIHTSGPLIEWLOQNRPEYIDLLRSVLKRGQVEIVVAGFY 85
 QY 76 TAILPLPLSRVEAQVQDRVKEELFE--VSPKGFWLP--LADPIIPAILKONGVEYLF 131
 Db 86 EPVLASIP--KEDRIEQIRLMKEWAKSIGFDARGVWLTERTVQWPELVKTLKESGDIYVI 142
 QY 132 ADEAMLSAHLNS-----AIKPKPLPHLIKAQREKRFYISYLLRLREL 175
 Db 143 VDDYHFMSAGLSKEELYWPTYTDEGGSEVIAVFP-----DEKRLYL--IPRPV 189
 QY 176 RKAIKLVF-----EGKVTILKVKDIEAVPVAVNTAVMLIGRLPLMNPCKKVASWIEDKN 229
 Db 190 DKVLEYLHSLIDDESKVAVFHDDGKFGIWP-----TYEWYIEKG 231
 QY 230 ILLYGTDFIEGYRDIAGRMSVGLLEVIDELNSLCPSEL-KHSGREL-----YLRT 281
 Db 232 WL-----REFFDRISSD-----EKNIMLYTEYLEKYPKRGVLVPLPIASYFEM 274
 QY 282 SSWA--DKSLRIWREDEGNARLNNLYN-----MRGEL-----AF 313
 Db 275 SEWSLPAKQARLEVEFVNELKVGIFEYRVFVGIGIWKNFYKYPESNYMHKRLMWSK 334
 QY 314 LAENSDARGWPLPERRLDAFRAIYN--WRG 342
 Db 335 LVNRRN-----PEARKYLLRAQCNDAYWHG 358

RESULT 3
 US-07-894-212A-2
 ; Sequence 2, Application US/07894212A
 ; Patent No. 5366883
 ; GENERAL INFORMATION:
 ; APPLICANT: ASADA, KIYOZO
 ; APPLICANT: UEMORI, TAKASHI
 ; APPLICANT: MUKAI, HIROYUKI
 ; APPLICANT: KATO, IKUNOSHIN
 ; APPLICANT: LADERMAN, KENNETH
 ; APPLICANT: ANFINSEN, CHRISTIAN
 ; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CUSHMAN DARB & CUSHMAN
 ; STREET: 1100 NEW YORK AVENUE, N.W.
 ; CITY: WASHINGTON, D.C.
 ; COUNTRY: U.S.A.
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/894,212A
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KOKULIS, PAUL N.
 ; REGISTRATION NUMBER: 16773
 ; REFERENCE/DOCKET NUMBER: 95469/C-1195
 ; TELEPHONE: (202) 861-3000
 ; TELEFAX: (202) 822-0944
 ; TELEFAX: 6714627 CUSH
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 649 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear

```

;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 650 amino acids
;

```

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0453 US
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 3:
LENGTH: 301 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: OVARUT02
CLONE: 2595635
US-09-002-298-3

Query Match 4.8%; Score 85.5; DB 3; Length 301;
Best Local Similarity 25.0%; Pred. No. 0.91;
Matches 30; Conservative 17; Mismatches 56; Indels 17; Gaps 3;
QY 184 EGVTLKVKDIEAVPVAVNTAVMLIGRLPLMNPKKVASWTEDKNILLYGTDTIEFIGYR 243
Db 171 ESKIRVHKEMGIIPSWGTTLEIIGSRQALKVLGSKALDKSNALNIGMVEEVLOSS 230
QY 244 DIAGRAMS-----VEGLLEVIDELNSLCPSELKHSGRELYLRTSSWADKSL--RIW 292
Db 231 DETKSLEAEQWLKQFIQGPPEVIRALKSVC-----SGRELYLEALQNERDLLGTW 284

RESULT 6
US-08-773-870-4
Sequence 4, Application US/08773870
Patent No. 5912143
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Goll, Surya K.
TITLE OF INVENTION: NOVEL HUMAN MAGE-LIKE PROTEIN
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/773,870
FILING DATE: Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0179 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 369 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

LIBRARY: GenBank
CLONE: 533511
US-08-773-870-4

Query Match 4.6%; Score 82.5; DB 2; Length 369;
Best Local Similarity 19.7%; Pred. No. 2.6;
Matches 67; Conservative 57; Mismatches 95; Indels 121; Gaps 18;
QY 11 QYAEIPKSEPKAVIEKAYIPVIE-----TLKEEPFGLNITGYTLKFLPKDILVKGGI-- 63
Db 85 QSAQIACSSPSV--ASPLDQSDGSSSQKEE-----SPSTLQVLPDSESLPRSEIDE 136
QY 64 -ASDLIEITGTSYTAILPLPLSRVEAQVORDRVKELEFVSPKGFVLPDELADPIIPAIL 122
Db 137 KVTDLVQFLFKY-----OMKEPITKAILLESVTKNY----- 168
QY 123 KDNQYELF--ADEAMLFSAHLNSAIKPIKPLPHLIKAQREKRFYISYLLRELKRAIK 180
Db 169 -EDHFLPLFSEASECMLLVFGID--VKEVDPTGH-----SFLVITSL----G 208
QY 181 LVPEGKVTLVKQLEAVP---VWVAVNTAVMLIGRLPLMNPKKVASWIEDKNILLYGTDI 237
Db 209 LTDGMLS---DVQSMPTGILILILSIIFIEG---YCTPEEV----- 245
QY 238 EFIGYRDIAGRMSVEGLEVIDELNSLCPSELKHSGRELYLRTSSWADKSLRIWREDEG 297
Db 246 -----IWEALNMGLYDGMHL-----IYGEPRKLLTQDVQVQENYLEYRQVPG 288
QY 298 N--ARLNMLYNMRGE-----LAFLE--NSDARGWPL 325
Db 289 SDPARVEFLWGPRAHAEIRKMSLLKFLAKVNGSDPRSFP 328

RESULT 7
US-08-881-706-2
Sequence 2, Application US/08881706
Patent No. 6245969
GENERAL INFORMATION:
APPLICANT: Chong, Joane
APPLICANT: Li, Jianming
TITLE OF INVENTION: Receptor Kinase BIN1
FILE REFERENCE: 07251/022001
CURRENT APPLICATION NUMBER: US/08/881,706
CURRENT FILING DATE: 1997-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 1196
TYPE: PRT
ORGANISM: Arabidopsis sp.
US-08-881-706-2
Query Match 4.6%; Score 82.5; DB 4; Length 1196;
Best Local Similarity 20.3%; Pred. No. 16;
Matches 77; Conservative 56; Mismatches 139; Indels 107; Gaps 17;
QY 42 LMITGYTLKFLPKDILVKGGIASDLIEITGYTAT-----LPLPLS 85
Db 153 LNVSSNTLDFPK-----VSGGLKLSLEVLDLSANSISGANVGVWVLSGCGELKHLAIS 208
QY 86 --RVEAQVORDR--VKELEFVSPKGF--WLPDELADPIIPAILKDNQVYELFAD----- 133
Db 209 GNKISGDDVDSRCVNLLEFLDVSSNNFSTGIFPLGDCALQHLDISGNK--LSGDFSRAIST 267
QY 134 --EAMLFSAHLNSAIKPIKPLPHLIKAQREKRFYISYLLRELKRAIKLVFEGKV---- 187
Db 268 CTELKLLNISSNQFVGPIPLP-----LKSLOYLSLAENK-----FTGEIPDFL 311
QY 188 -----TLKVKDIE-----AVPVWVAVNTAVMLIGRLPLMNPKKVASWIEDKNILLYGTD 236
Db 312 SGACDTLTGLDLSGNHFFYGVPPFFG---SCSLLESLSLSSNFSGELPMDTLKMRGLK 369

QY 237 IEFYGTAGRMVSEGLLEVIDEINSELCPSEL-----KHSGRELYL 279
 Db 369 VLDLSFNFSGE-----LPESLTNLSASLLTDLSSNFSGPILPNLQCNPKNTLQELYL 423
 QY 280 RFSWADKSLRTWREDEGNARLNMLYN-----MRGELAFLENSDARGW-----PLPE- 327
 Db 424 QNNGFTGKIPPTLPSNCSSELVSLHSFNLSGTIPSSLSGLSKLRDLKLWLNWLGEEIPQE 483
 QY 328 ----RRLDAFRAIYNDRG 342
 Db 484 LMVXTLETLLIDFNDLTG 502

RESULT 8
 US-09-141-135-2
 ; Sequence 2, Application US/09141135
 ; Patent No. 5981729
 ; GENERAL INFORMATION:
 ; APPLICANT: CHUN, Jong Yoon
 ; APPLICANT: LEE, Yong Hun
 ; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Absciscic
 ; TITLE OF INVENTION: Acid Isolated from Arabidopsis thaliana
 ; FILE REFERENCE: 1942/31
 ; CURRENT APPLICATION NUMBER: US/09/141,135
 ; CURRENT FILING DATE: 1998-08-27
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: WordPerfect 6.1/Windows
 ; SEQ ID NO 2
 ; LENGTH: 235
 ; TYPE: PRT
 ; ORGANISM: Arabidopsis thaliana
 US-09-141-135-2

Query Match 4.6%; Score 82; DB 2; Length 235;
 Best Local Similarity 20.1%; Pred. No. 1.4;
 Matches 45; Conservative 37; Mismatches 82; Indels 60; Gaps 10;
 QY 157 KAQREKFRYISYLLRELRLKAIKLVFEKGVTLKVKDIEAVPVWVAVNTAVMLGRPLM 216
 Db 27 KSNQKRFN-----EEQIKSLLEIFSETRLEPR-----KKVQVARELGL 66
 QY 217 NPKKVASWEDKNILLYGTDIE-----FIGYDIAGRMV-----EGLEVIDELNSFL 265
 Db 67 QPROMTIWFQNKRAWKTKOLEKEYNTLRANYNLTASQFEIMKKEKQSLVSELQRLNEEM 126
 QY 266 -CPSELKH-----SGRELYLRTSSWADKSLRIWREDEGNARLN-----MLNNMR 308
 Db 127 QRKEKHHKCCGCGDLALSSSTESHNGKSEPEGRDQGSVLCNDGDYNNNIKTEYFRVO 186
 QY 309 G-----ELAFLENSD-----ARGWP--LPERRLDAFRAIYNDR 340
 Db 187 GETDHELMNIVEKADDSCLTSSNNGWGFNSDLSLLDQSSSNYPNW 230

RESULT 9
 US-09-286-805-4
 ; Sequence 4, Application US/09286805
 ; Patent No. 6117990
 ; GENERAL INFORMATION:
 ; APPLICANT: Bonini, James A.
 ; APPLICANT: Borowsky, Beth E.
 ; TITLE OF INVENTION: DNA Encoding Orphan SNORF1 Receptor
 ; FILE REFERENCE: 58987
 ; CURRENT APPLICATION NUMBER: US/09/286,805
 ; CURRENT FILING DATE: 1999-04-06
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: PatentIn Ver. 2.0 - beta
 ; SEQ ID NO 4
 ; LENGTH: 338
 ; TYPE: PRT
 ; ORGANISM: Rattus norvegicus

US-09-286-805-4

Query Match 4.6%; Score 82; DB 3; Length 338;
 Best Local Similarity 29.1%; Pred. No. 2.5;
 Matches 30; Conservative 17; Mismatches 34; Indels 22; Gaps 5;

QY 88 EAQVORVRKEELFEVSPKGF---WLPDLADPIIPAILKONGVEYLFADKAMLFSAHLNS 144
 Db 237 ERVARRERKAATLGIAAFAFLVSWLPYIIDAVIDAYMNFITPAYVY--BILVWCYVYNS 294
 QY 145 AIKPIKPLPHLIKAEKREFRYISYLLRELRLKAIKLVFEKGV 187
 Db 295 AMNP-----LIYA-----FFYPWF-----RRAIKLIVSGKV 320

RESULT 10
 US-08-231-342-23
 ; Sequence 23, Application US/08231342
 ; Patent No. 5827684
 ; GENERAL INFORMATION:
 ; APPLICANT: Sreekrishna, Kotikanyadanam
 ; APPLICANT: Prevatt, William D
 ; APPLICANT: Thill, Gregory P
 ; APPLICANT: Davis, Geneva R
 ; APPLICANT: Koutz, Patricia
 ; APPLICANT: Barr, Kathryn A
 ; APPLICANT: Hopkins, Sharon A
 ; TITLE OF INVENTION: Production of Bacillus Entomotoxins in
 ; TITLE OF INVENTION: Methylophilic Yeast
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fitch, Even, Tabin & Flannery
 ; STREET: 135 S. LaSalle St.
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: U.S.A.
 ; ZIP: 60603-4277
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/231,342
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/926,448
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Feder, Scott B
 ; REGISTRATION NUMBER: 33,129
 ; REFERENCE/DOCKET NUMBER: 52627
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 312-372-7842
 ; TELEFAX: 312-372-7848
 ; INFORMATION FOR SEQ ID NO: 23:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 448 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-231-342-23

Query Match 4.6%; Score 82; DB 2; Length 448;
 Best Local Similarity 22.6%; Pred. No. 3.9;
 Matches 67; Conservative 36; Mismatches 89; Indels 104; Gaps 16;

QY 26 AYIEVI-ETLIKE---EPEGLNITGYTLKFLPKDII-LVKGGIASDLIEIGTSTYALP 80
 Db 199 AAIPQLPQTSLENIPEPTSLDDSG-----VLPKDAVRVKG-----SALLP 240


```

; APPLICATION NUMBER: US 746,121
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-977-434-4

Query Match 4.6%; Score 82; DB 1; Length 893;
Best Local Similarity 22.4%; Pred. No. 11;
Matches 85; Conservative 49; Mismatches 103; Indels 142; Gaps 22;

QY 19 EKVTIEKAVIP-----VIETL--LKEPPFGNITGYTLKPLPKDIIIVKGGI 63
Db 344 KPA--EAYIPLHHRNAQNLEKEVLKLELLEDP-GAKIVGQNLKFDYK-VLMVKG-- 397
QY 64 ASDLIBIIGTSYTAILPLPLSRVEAQVQRVRKKEELFEVSPKGFMLPELADPIIPALK 123
Db 398 -----VEPVPPYEDT-----MIAAYLLEPNKKNLDDLA-----LK 429
QY 124 DNGEYLFDAEAMLSAHL---NSAIKPKPLPHLIKAQREKFRY-----ISYLLRE 174
Db 430 FLGYKMTSYOELMSFPLGFGFADVPV-----EKAANYCEDADITRLYKT 478
QY 175 LRKAIRLVEGKVTLLKVKDIEAVPVVAVNTAVMLIGRLPLMNPKKVASWIEDKNILL-- 232
Db 479 L-----SLKLHEADLENVYKI-----EMPLYN---VLARMELNGVYVD 515
QY 233 -----YGTDFIEFG---YRDIAG-----RMSVEGL----- 254
Db 516 EFLKLLSEYGGKLEELAEIYR-IAEPFNINSKQVSRILFEKLGIKPRGKTKTKGDY 574
QY 255 ---LEVIDEL--NSELCPSELKHSRGELVYRTSSWADKSLRIWREDEGNARINMLNMG 309
Db 575 STRIEVLEELAGHEHILILEY--RKOKLKSTYIDALPKVWNPKTG--RIHASFNTQG 630
QY 310 ELAFLAENSARGWPLPER 328
Db 631 TATGRLLSSDPNQLNLPK 649

RESULT 13
US-08-458-819-4
; Sequence 4, Application US/08458819
; Patent No. 5795762
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF

```

```

; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; ZIP: 07110-1199
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: Macintosh
; OPERATING SYSTEM: 7
; SOFTWARE: Wordperfect 2.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,819
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/977,434
; FILING DATE: 23-FEB-1993
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441
; FILING DATE: 12-JAN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 063,509
; FILING DATE: 17-JUN-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 899,241
; FILING DATE: 22-AUG-1986
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US90/07641
; FILING DATE: 21-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 585,471
; FILING DATE: 20-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 455,611
; FILING DATE: 22-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 609,157
; FILING DATE: 02-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 557,517
; FILING DATE: 24-JUL-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Luann Cserr
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: Case No. 5795762 8753
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2972
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 893 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-458-819-4

```

Query Match 4.6%; Score 82; DB 1; Length 893;
Best Local Similarity 22.4%; Pred. No. 11;
Matches 85; Conservative 49; Mismatches 103; Indels 142; Gaps 22;

QY 19 EPKVIKAYIP-----VIETL--IKEEPFGLNITGYTLKFLPKDIIIVKGGI 63
Db 344 KPR--EAYIPLHHRNAQNLDKEVLKLEKLEIDP-GAKIVGQNLKFDYK-VLMVKG-- 397
QY 64 ASDLIEIGTSYTAILPLPLSRVEAQVDRVKEELFEVSPKGFWLPDLADPIIPAILK 123
Db 398 -----VEPVPPYFDT-----MIAAYLLEPNKKFNLDLLA-----LK 429
QY 124 DNGEYLFADAEAMLSAHL--NSAIKPIKPLHLIKAQREKFRY-----ISYLLRE 174
Db 430 FLGYKMTSYQELMSFSPLFGSFADVPV-----EKAANYSCEDADITYRLYKT 478
QY 175 LRKAIKLVFEGKVTLLKVKDIEAVPVVAVNTAVMLIGRLPLMNPKNKVASWIEDKNILL-- 232
Db 479 L-----SLKLHEADLENVFIKI-----EMPLVN-----VLARMELNGVYVDT 515
QY 233 -----YGTDIETFG---YRDIAG-----RMSVEGL----- 254
Db 516 EFLKLLSEYGGKLEELAEIYR-IAGEPFNINSKQVSRILFEKLGKIPRGKTKTKGDY 574
QY 255 ---LEVIDEL--NSELCPSELKHSRELYLRTSSWADKSLRIWREDEGNARLNMLNMRG 309
Db 575 STRIEVLEELAGEHEIIPILLEY--RKIOKLKSTYIDALPKMNPKNKTG--RIHASFNOTG 630
QY 310 ELAFLAENSARGWPLPER 328
Db 631 TATGRLLSSDPNQLNLPK 649

RESULT 14

US-09-105-697-10
; Sequence 10, Application US/09105697
; Patent No. 6228628
; GENERAL INFORMATION:
; APPLICANT: Gelfand Ph.D., David H.
; APPLICANT: Reichert, Fred L.
; TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Roche Molecular Systems
; STREET: 1080 U.S. Highway 202
; CITY: Branchburg
; STATE: New Jersey
; COUNTRY: United States
; ZIP: 08876
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/105,697
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry Ph.D., Douglas A.
; REGISTRATION NUMBER: 35321
; REFERENCE/DOCKET NUMBER: 1043
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510)814-2974
; TELEFAX: (510)814-2977
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 893 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-105-697-10

Query Match 4.6%; Score 82; DB 4; Length 893;
Best Local Similarity 22.4%; Pred. No. 11;
Matches 85; Conservative 49; Mismatches 103; Indels 142; Gaps 22;

QY 19 EPKVIKAYIP-----VIETL--IKEEPFGLNITGYTLKFLPKDIIIVKGGI 63
Db 344 KPR--EAYIPLHHRNAQNLDKEVLKLEKLEIDP-GAKIVGQNLKFDYK-VLMVKG-- 397
QY 64 ASDLIEIGTSYTAILPLPLSRVEAQVDRVKEELFEVSPKGFWLPDLADPIIPAILK 123
Db 398 -----VEPVPPYFDT-----MIAAYLLEPNKKFNLDLLA-----LK 429
QY 124 DNGEYLFADAEAMLSAHL--NSAIKPIKPLHLIKAQREKFRY-----ISYLLRE 174
Db 430 FLGYKMTSYQELMSFSPLFGSFADVPV-----EKAANYSCEDADITYRLYKT 478
QY 175 LRKAIKLVFEGKVTLLKVKDIEAVPVVAVNTAVMLIGRLPLMNPKNKVASWIEDKNILL-- 232
Db 479 L-----SLKLHEADLENVFIKI-----EMPLVN-----VLARMELNGVYVDT 515
QY 233 -----YGTDIETFG---YRDIAG-----RMSVEGL----- 254
Db 516 EFLKLLSEYGGKLEELAEIYR-IAGEPFNINSKQVSRILFEKLGKIPRGKTKTKGDY 574
QY 255 ---LEVIDEL--NSELCPSELKHSRELYLRTSSWADKSLRIWREDEGNARLNMLNMRG 309
Db 575 STRIEVLEELAGEHEIIPILLEY--RKIOKLKSTYIDALPKMNPKNKTG--RIHASFNOTG 630
QY 310 ELAFLAENSARGWPLPER 328
Db 631 TATGRLLSSDPNQLNLPK 649

RESULT 15

PCT-US91-07035-4
; Sequence 4, Application PC/TUS9107035
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Abramson, Richard D.
; TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
; TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cetus Corporation
; STREET: 1400 Fifty-third Street
; CITY: Emeryville
; STATE: California
; ZIP: 94608
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/07035
; FILING DATE: 19910930
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,490
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,466
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 590,213
; FILING DATE: 28-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 523,394
; FILING DATE: 15-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 143,441

' Thu Apr 4 16:34:56 2002

Search completed: April 4, 2002, 14:40:44
Job time: 193 sec

FILING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 063,509
FILING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FILING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Sias Ph.D, Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: Case No. 2580
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-420-3300
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 893 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US91-07035-4

Query Match 4.6%; Score 82; DB 5; Length 893;
Best Local Similarity 22.4%; Pred. No. 11;
Matches 85; Conservative 49; Mismatches 103; Indels 142; Gaps 22;

QY	19	EPKVIKAYIP-----VETL--IKPEPGLNITGYTLKPLPKDIIILVKGGI	63
Db	344	KPK--EAYIPLHRRNAQNLEKEVLKLEIDDP-GAKIVGQNLKFDYK-VLMVKG--	397
QY	64	ASDLIEICTSYTALPLPLSRVQAQVORDRVKEELFEVSPKGFWLPDLADPIIPAIK	123
Db	398	-----VEPVPPYEDT-----MIAAYLLEPNEKKFNLDL-----LK	429
QY	124	DNGYEYLFDAEAMLSAHL---NSATKPKPLHLKAQREKFRY-----ISYLLRE	174
Db	430	FLGYKMTSYQELMSFSPLFGFSADVPV-----EKAANYSCEDADITYRLYKT	478
QY	175	LRAIKLVPEGKVTLVAKDIEAVPVVAVNTAVMLIGRLPLMNPKNVSWIEDKNILL--	232
Db	479	L-----SLKLHEADLENVEFYK-----EMPLVN-----VLARMELNGVYVDT	515
QY	233	-----YGTDFIEFIG---YRDIAG-----RMSVEGL-----	254
Db	516	EFLKLESEYGGKLEELAEIYR-IAEPFNINSKQVSRILLFEKIGIKPRGKTKTKGDY	574
QY	255	---LEVDEL--NSELCPSKELHSGRELYRTSSWADKSLRIWREDEGNARLNMLYNMRG	309
Db	575	STRIVLEELAGEHEIPLILEY--RKIQKLSYIDALPKMVNPKTG--RIHASFNOTG	630
QY	310	ELAFIENSARGWPLPER	328
Db	631	TATGRSSSDPNLQNLPK	649

Thu Apr 4 16:34:56 2002

us-09-407-806a-4_1.rai

Page 10

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:38:16 ; Search time 21.67 seconds
(without alignments)
1216.262 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 1778
Sequence: 1 LRALVFHGNLQYAEIPKSEP.....ERRLDAPRAIYNDWRNGNEP 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1229	69.1	364	2 E71144	hypothetical prote
2	151	8.5	686	1 ALDYAT	amylase A (EC 3.2.
3	126.5	7.1	633	2 G71241	probable alpha-am
4	120.5	6.8	655	2 E75206	alpha-amylase (or
5	116.5	6.6	649	2 A49512	alpha-amylase (EC
6	111	6.2	4725	1 A44357	dyein heavy chain
7	110	6.2	1362	2 A75207	amylopullulanase p
8	106.5	6.0	392	2 H82815	fibrillar assembly
9	106.5	6.0	467	2 B64501	alpha-amylase (EC
10	104.5	5.9	1251	2 JH0256	botulinum neurotox
11	102	5.7	738	2 S52737	NADH dehydrogenase
12	102	5.7	853	2 T08162	amylopullulanase (
13	101.5	5.7	705	2 T48464	hypothetical prote
14	101.5	5.7	1396	2 F81686	DNA-directed RNA p
15	101	5.7	394	2 F75006	hypothetical prote
16	100	5.6	336	2 H71103	hypothetical prote
17	99.5	5.6	630	2 B64514	hypothetical prote
18	99	5.6	391	2 E64474	hypothetical prote
19	98	5.5	1430	2 D82533	RNA polymerase bet
20	97.5	5.5	1396	2 G71529	DNA-directed RNA p
21	97.5	5.5	1473	2 T38791	probable ferredoxi
22	97	5.5	480	2 E72682	hypothetical prote
23	97	5.5	924	1 JC1369	DNA-directed DNA p
24	96.5	5.4	819	2 E70105	p115 protein homol
25	96.5	5.4	1367	2 C72360	DNA polymerase III
26	96	5.4	1210	2 E64979	hypothetical 138.1
27	96	5.4	1242	2 S51246	probable DNA repla
28	95.5	5.4	588	2 T37542	conserved hypothet
29	95.5	5.4	684	2 T47694	probable serine/th

ALIGNMENTS

RESULT 1

E71144
hypothetical protein PH0368 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000

C:Accession: E71144

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sa

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: E71144

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

A:Residues: 1-364 <KAW>

A:CROSS-references: GB:AP000002; NID:G3236129; PIDN:BAA29442.1; PID:d1030385; PID:g3

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenB

C:Genetics:

A:Gene: PH0368

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0368

Query Match 69.1%; Score 1229; DB 2; Length 364;
Best Local Similarity 72.2%; Pred. No. 4.5e-86;
Matches 262; Conservative 47; Mismatches 36; Indels 18; Gaps 18;

QY 1 LRALVFHGNLQYAEIPKSE-PKVIKAYIPVIETLIKEE-PFGLNITGYTLKFLPKDII- 57

Db 1 MRALIFHGNLQYAEIPKSEISKVIEKSYFPTISELIKREIPFGLNITGYSLQFLPQELIH- 60

QY 58 LVKGGIADSLIELIGTSYT-ATLPLPLSRVEAQVQRD-VKEELFEVSPKGFWLPDLA- 114

Db 61 LIKEGIESLEILGTSYTHAILPLTLRSIAQKRDREIKREIFEVSPGGFWLPDLAY 120

QY 115 DPTIIPAILKDNQYEVLFAD-EAMLSAHLNSAIKPKPL-PHLIKAQREKRPYSYLL- 171

Db 121 DPTIIPAILKDNQYEVLFADGEAMLSAHLNSAIKPKPLPYLYLKAQGEVGYLYNLLG 180

QY 172 LRELKRAIKLVFEGKVTLK-VKDIKAVPVVAVNTAVML-IGRLPLMPKPKVASIEDKN 229

Db 181 LRELKKAINTLFGGKVTLEAVKDIEAIPVWSINIAIMLGACRFPLMSPKRVANNIKGD 240

QY 230 -ILLVGTDTIEFTGYRDIAG-RMSVEGLEVIDELSELG-PSELKHSRELILRTSSWA- 285

Db 241 EILLVGTDTIEFTGYRSIAGHKAISGLMEVINELOGECLPRDRHNGRLRLRTSSWAP 300

QY 286 DKSRLRWDEGNARLNML-YNMRELAFLAENSADRGW-PLPERRLDFAFRIYNDWRG- 342

Db 301 DKSRLRWKDEGNARLNMLTYCMDGEFAFLAENSADRGWELPERRLDFAKAIYKWRNE 360

QY 343 NGE 345

Db 361 NGK 363

RESULT 2

ALDYAT

amylase A (EC 3.2.1.-) - Dictyoglomus thermophilum

C:Species: Dictyoglomus thermophilum

C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 18-Jun-1999

C:Accession: S00628; A34969

R:Fukushima, S.; Kanizono, A.; Horinouchi, S.; Beppu, T.

Eur. J. Biochem. 174, 15-21, 1988

A:Title: Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic

A:Reference number: S00628; MUID:88225097

A:Accession: S00628

A:Molecule type: DNA

A:Residues: 1-686 <FUK>

A:Cross-references: EMBL:X07896; NID:g2688; PIDN:CAA30735.1; PID:g2689

A:Accession: A34969

A:Molecule type: protein

A:Residues: 2-13 <FUK2>

C:Genetics:

A:Gene: amyA

C:Superfamily: Dictyoglomus thermophilum amylase A

C:Keywords: glycosidase; hydrolase

F:2-686/Product: amylase A #status experimental <MAT>

Query Match 8.5%; Score 151; DB 1; Length 686;

Best Local Similarity 21.6%; Pred. No. 0.0011;

Matches 91; Conservative 63; Mismatches 111; Indels 156; Gaps 24;

QY 8 GNQVAETPKSPKVEKA----YTPVETLIKEPPFGLNT--TGYTLKFLPKD----- 55

Db 18 GNFD-----VIERAYEMSKPLINFFPKHPDPFPIVHFSGLLLWLEKNHPEVE 68

QY 56 ---ILLVKGGIASDLIETIGTSYTAILPLPLSRVEAQVDRVKE-----ELFEV 103

Db 69 KLIKMAERGQTEF-----VSGGFYEPILPIP-----DKDKVQOIKKLNKYIYDFEQG 116

QY 104 SPKGFWLPE-LADPIIPAILKNDGYEYLFADAMLFSAHLSAIAKPIKPLHLIKAQREK 162

Db 117 TPKGMWLAERWEPHLVRYIAEAGIEYVVDDAHFFSVGL-----K 157

QY 163 RFRYSYLLRLRELKAIKL-VFEGKVTYLVKVDIEAVPVAVTAVMLIGRLPLMNPVKV 221

Db 158 BEDLFGYILMEE--QGYKLVAF--PISMKRLY-----IPFADPEET 195

QY 222 ASWI-----EDKN--ILLYGTDIEFGYRDIAGRMSVEGLE-----VIDELNS 263

Db 196 IYLDKFASDKSKTALLFDGGEKFGLPDITYRVYEEGWLTFVSKIFENFLVTPVNL 255

QY 264 ELCPSELKHSGRELYLRTSSW-----ADKSLR----- 290

Db 256 YTYMQRVKPKGR-IYLPATSYREMMEWVLPPEAQKELEELVEKLKTENLWDKTSPPYVKG 314

QY 291 LWRE-----DSGN-ARLNMLYNNMGEALFAENSDARGWLPERRL--DAFRAYND--WR 341

Db 315 FWRNPLAKYDESNHMKMLYVWK-----KVQDSPEEVEYKAMEEVEFQGGANDAYWH 367

QY 342 G 342

Db 368 G 368

RESULT 3

G71241

probable alpha-amylase - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: G71241

R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki

M.; Onfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: G71241

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-633 <RAW>

A:Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29262.1; PID:g3256579

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0193

C:Superfamily: Dictyoglomus thermophilum amylase A

Query Match 7.1%; Score 126.5; DB 2; Length 633;

Best Local Similarity 22.3%; Pred. No. 0.075;

Matches 87; Conservative 59; Mismatches 121; Indels 123; Gaps 24;

QY 21 KVIEKAVIPVETLIKEEP--FGLNITGYTLAKFL----PKDILLVKGGTASDLIE-IIG 72

Db 25 RAYEKAYRPFLEVL-EYYPNMKVAVHLSGVLVLEWLEERNRPEYIDLLKSLIKKGQVELVWA 83

QY 73 TSYTAILPLLP-LSRVEAQVDRVKEELFEVSPKFWLPE-LADPIIPAILKNDGYEYL 130

Db 84 GFVEPILVAIPEEDRVE-QIKLSKGWARKMGYEARGLWLTERTVWPEPELVKTLREAGIEYV 142

QY 131 FADLAMLFSAHLS-----AIPKIPKPLHLIKAQREKFRFYSYLLLR 174

Db 143 ILDDYHMSAGLSKEELFWPYTTENGEATVPEPI-----DEKLRYL--IPFRP 189

QY 175 LRKAIKLVF-----EGKVTYLVKVDIEAVPVAVTAVMLIGRLPLMNPVKVSVI--- 225

Db 190 VNETLEVHLSLADESKVAVFHDDGKFCAPWGTHELIVYERG-----WLKEF 237

QY 226 -----BDK-NILLYGTDIEFG-----YRDIAG--RMSVEGL-----LEV 257

Db 238 FDRISDDKINMLYS---EYLSKFRPKGLVLYPIASIFYEMSEWSLPARQAKLFFFEFIKK 294

QY 258 IDLSELCPSELKHSGRELYLRTSSWADKSLRWDEGN---ARLNMLYNNMGEALFL 314

Db 295 LKELN-----LFPEKYRIFVRGGVWKNN---FLYKYPEGNTMHRKMLMSKL-----L 337

QY 315 AENSARGMWPLPERRLDRAFRAYND--WRG 342

Db 338 RNN-----PTARIFVLRACNDAYWHG 359

RESULT 4

E75206

alpha-amylase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Ors

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: E75206

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: E75206

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-655 <KAW>

A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49100.1; PID:g545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: amyA; PAB0118

C:Superfamily: Dictyoglomus thermophilum amylase A

Query Match 6.8%; Score 120.5; DB 2; Length 655;

Best Local Similarity 20.1%; Pred. No. 0.22;

Matches 79; Conservative 62; Mismatches 117; Indels 135; Gaps 20;

```

QY 24 EKAYIPVETLKEEP---FGLNITGYTLKFL---PKDILLVKGGIASDLIELIGTS-Y 75
Db 28 EKAYRPFLE-IIEEYPMKVAIHISGLVLEENKPYIDLLKSLVRKGGQVEIVVAGFY 86
QY 76 TAILPLPLSRVAVQVDRVKE-----ELFEVSPKGEWLPE-LADPIIPAILKONG 126
Db 87 EPVLAAP-----KEDRLQIYLLKEWAKKIGYDAKGLWLTERTVWQPELVKTLREAG 138
QY 127 YEYLFADAMLSAHLNS-----AIKPIKPLHLIKAKQREKFRFYISYL 170
Db 139 IEYVVVDDYHFMASGLSKDQLEWPPYTTEDGGEVITVFPI-----DEKRLYL--I 185
QY 171 LLRELKAIKLVF-----EGKVTLLKVKDIEAVPVVAVNTAVMLIGRLPLMNPCKVASW 224
Db 186 PFRPVDKVIYSLHSLASESKVAVFHDDGKFGIW-----PM-----TYEW 227
QY 225 IEDK-----NILLXGTDLIE-----FIGYRDIA-----GRM 249
Db 228 VYEGWLREFFDVRSSDEAINIMLYSEYLOKFKPKGLVLYPLTASYFEMSEWSLPAQOAKL 287
QY 250 SVEGLLEVIDELNSELCPSELKHSRGRELYLRTSSWADKSLRTWRDEGNARLNMLYNRG 309
Db 288 FVE-FVEKLKELN-----MFERYRVFVGGIWKNFYKYPEANYMHKRMMLSL-- 336
QY 310 ELAFLAENSADRGWPLPERRLDAFRAYNDWRG 342
Db 337 ----LRDNP SARFVLRAOCNDAY-----WHG 359

RESULT 5
A49512
alpha-amylase (EC 3.2.1.1) - Pyrococcus furiosus
C:Species: Pyrococcus furiosus
C>Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C:Accession: A49512
J:Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen, C.B.
J. Biol. Chem. 268, 24402-24407, 1993
A:Title: alpha-amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus.
A:Reference number: A49512; MUID:94043280
A:Accession: A49512
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-649 <LAD>
A:Cross-references: GB:L22346; NID:q347939; PIDN:AAA72035.1; PID:q347940
C:Genetics:
A:Gene: amyA
A:Start codon: GTG
C:Function:
A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A:Pathway: glycogen/starch degradation
C:Superfamily: Dictyoglomus thermophilum amylase A
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 6.6%; Score 116.5; DB 2; Length 649;
Best Local Similarity 20.2%; Pred. No. 0.45;
Matches 79; Conservative 64; Mismatches 117; Indels 131; Gaps 21;

QY 24 EKAYIPVETLKEEP---FGLNITGYTLKFL---PKDILLVKGGIASDLIELIGTS-Y 75
Db 29 EKCYWPFLETL-EYPMKVAIHTSGPIELWLDQNRPEYIDLLSLVRKGGQVEIVVAGFY 87
QY 76 TAILPLPLSRVAVQVDRVKEPEF---VSPKGEWLPE-LADPIIPAILKONGYEYLF 131
Db 88 EPVLASIP---KEDRIEQIRLMKEWAKSIGFDARGVWLTERTVWQPELVKTLKESGDYVI 144
QY 132 ADEAMLSAHLNS-----AIKPIKPLHLIKAKQREKFRFYISYLLREL 175
Db 145 VDDYHFMASGLSKDELYWPPYTTEDGGEVIAVFPI-----DEKRLYL--IPFRPV 191
QY 176 RKAIKLVF-----EGKVTLLKVKDIEAVPVVAVNTAVMLIGRLPLMNPCKVASWIEDKN 229
Db 192 DKVLEYLHSLDGDSEKVAVFHDDGKFGIWP-----TYEWYVERG 233

```

```

QY 230 ILLYGTDFIEGYRDIAGRMVSVEGLLEVIDELNSELCPSEL-KHSGREL-----YLRT 281
Db 234 WL-----REFDRISSD-----EKINLMLEYLEYKYPKGLVYLPPIASFEM 276
QY 282 SSWA--DKSLRWREDEGNARLNMLYN-----MRGEL-----AF 313
Db 277 SEWSLPKARQLRVEFVFNELKVGKIFERYRVFVGGIWKNFYKYPESNYMHKRMMLVSK 336
QY 314 LAENSDARGWPLPERRLDAFRAYND--WRG 342
Db 337 LVRNN-----PEARKYLLRAOCNDAYWHG 360

RESULT 6
A44357
dynein heavy chain, cytosolic - slime mold (Dictyostelium discoideum)
N:Contains: dynein ATPase (EC 3.6.1.33)
C:Species: Dictyostelium discoideum
C>Date: 03-Feb-1994 #sequence_revision 02-May-1994 #text_change 19-Jan-2001
C:Accession: A44357; S28504
R:Koonce, M.P.; Grissom, P.M.; McIntosh, J.R.
J. Cell Biol. 119, 1597-1604, 1992
A:Title: Dynein from Dictyostelium: primary structure comparisons between a cytoplasmic
A:Reference number: A44357; MUID:93107159
A:Accession: A44357
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-4725 <KOO>
A:Cross-references: EMBL:Z15124; NID:g7226; PIDN:CAA78827.1; PID:g7227
A:Note: sequence extracted from NCBI backbone (NCBIP:121195)
C:Superfamily: dynein heavy chain, cytosolic
C:Keywords: ATP; heterotrimer; hydrolase; microtubule binding; nucleotide binding;
F:1969-1976/Region: nucleotide-binding motif A (P-loop)
F:2271-2278/Region: nucleotide-binding motif A (P-loop)
F:2669-2676/Region: nucleotide-binding motif A (P-loop)
F:3011-3018/Region: nucleotide-binding motif A (P-loop)
F:1975/Binding site: ATP (Lys) #status predicted
F:2277/Binding site: ATP (Lys) #status predicted
F:2675/Binding site: ATP (Lys) #status predicted
F:3017/Binding site: ATP (Lys) #status predicted

Query Match 6.2%; Score 111; DB 1; Length 4725;
Best Local Similarity 21.3%; Pred. No. 19;
Matches 80; Conservative 48; Mismatches 122; Indels 126; Gaps 17;

```

```

QY 12 YAEIPKSPKTEKAYIPVETLKEEPFGLNITGYTLKFLPKDILLVKGGIASDLIELI 71
Db 3358 YADLEKAEPAIIIEAQ--EAVSTIKKKH--LD-----EIKSLPKPPTPVK----- 3397
QY 72 GTSYTAILPLPLSRVAVQVDRVKEELFEVS-----PKGFWLPELADPIIPAILKONG 126
Db 3398 -LAMEAVCLMLGGKKLEWADIRKKIMEPNFITSINYDTKKMWTPKIREAITKGVLEDPG 3456
QY 127 YEYLFADAMLSAHLNSAIKPIKPLHLIKAKQREKFRFYISYL--LRELKAIKLVFE 184
Db 3457 FDY-----ETVNRASKACGGLVVKWATAQ-----TYTSEILDRKLRREEVEQLEN 3501
QY 185 GKVTLLKVKDIEAVPVVAVNTAVMLIGRLPLMNPCKVASWIEDKNILLYGTDIEFYGRD 244
Db 3502 AANELKRLKODEIVATITALE-----KSIATYKEEYATLIRET--EQIKTES 3545
QY 245 IAGRMVSVEGLLEVIDELNSELCPSELKHSRGRELYLRTSSWADKSLRWREDEGNARLNML 304
Db 3546 SKVKNKVDYSIALDLNLSN-----RGR-----NEQOSENFTQM- 3580
QY 305 YNMRGEL-----AFLA-----ENSDARGWPL-----PERRL 330
Db 3581 STVWGDVVLASAFIAYFGFDONFRDLMRKWMIRLDSVGIKFKSDLSVPSPFLSKPEERL 3640
QY 331 DAFRAYNDWRNGNEP 346

```

Db 3641 -----NWHANSLP 3648

RESULT 7

A75207
 amylopullulanase PAB0122 - Pyrococcus abyssi (strain Orsay)
 C:Species: Pyrococcus abyssi
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: A75207
 R:anonymous, Genoscope
 submitted to the EMBL Data Library, July 1999
 A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure
 A:Reference number: A75001
 A:Accession: A75207
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1362 <KAW>
 A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49104.1; PID:el51499
 A:Experimental source: strain Orsay
 C:Genetics:
 A:Gene: apu; PAB0122

Query Match 6.2%; Score 110; DB 2; Length 1362;
 Best Local Similarity 21.9%; Pred. No. 3.9;
 Matches 93; Conservative 61; Mismatches 150; Indels 120; Gaps 22;

QY 11 QVAETPKSPKV-----TEKAYIPV----- 30
 Db 174 KYSNLEPEQKIATVSEFTQDYIDLAVLFNLAWIDYIMNTPKALYDKVDFTGGYTR 233
 QY 31 --IETLIKKEPGLNITGTLKFLPKDIIILYKGGTASDLIEIGTSYTAILPLPL----- 84
 Db 234 KDVETVLKQWLLN---HTFEHEKINILLNG---NVEVTVPVYTH--PIGELNDF 284
 QY 85 ---SRVEAQVQR--DRVKEEL--FEVSPKGFWLPE--LADPIIPAILKDNQGYEYLFADFA 135
 Db 285 GWYEDFAQVKKANELYKEYLGAGKVTPKGGAESALNDKTL--ELIENSGKWKVMTDQL 343
 QY 136 MLFSAHLNSAI-----KPIKPLHLIKAQREKFRYSY-----LLREL 175
 Db 344 VLEKLGVPKTIIESYKPPWAQFGDKKILYLFPRNHDLSRVRGFRYAGMNOYDAVKNFVEEL 403
 QY 176 RKAILVFEGKVTAKVKDIEAVPVWVAVNTAVMLIGRLPLPNPKVASWIEDKNILLYGT 235
 Db 404 LKIQONDGSL--VIVITLDGENPW-----EHPDPGKLFLEELYRQLELQKGLIRTVT 458
 QY 236 DIEFTG-YRDIAGRM-----SVEGLL--EVIDELNSELCPSELKHSGRELY 278
 Db 459 PSEYTEMFGDKANKLTPKMKRDLDTTDDNVALLKAKTLGELYDMVGVTB-----EMQ 512
 QY 279 LRTSSWADKSLRIW-REDEGNARLNMLYMRGELAFIAENSD--ARGWPLPERRLDFAFRI 336
 Db 513 WPESWDGTLSTWIGEPQENIAYWLYLAR---KALFENKDNVKNWKAWEYL--FRAE 567
 QY 337 YNDW 340
 Db 568 GSDW 571

RESULT 8

H82815
 flmbrial assembly membrane protein Xf0369 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C:Accession: H82815
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequencing
 Nature 406, 151-157, 2000
 A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: H82815
 A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <SIM>

A:Cross-references: GB:AE003888; GB:AE003849; NID:g9105187; PIDN:AAF83179.1; GSPDB:GN

A:Experimental source: strain 9a5c

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer

as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La

chado, M.A.; Madeira, A.M.B.N.; Madeira, M.F.; Marino, C.L.; Marques, M.V.; Martins

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv

M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF0369

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

QY 27 YPVIETLIKKEEPF-GLNITGYTLKFLPKDIIILYKGGTASDLIEIGTSYTAILPLPL 85
 Db 27 YPVLKYPVVRGVFVL-----FSKQSVLVGVDSITAVK-----LQLS 67
 QY 86 RVEAQVQRDKVELFVSP-----KGFWLPELADPIIPAILKDNQGYEYLFADFA 137
 Db 68 R-----SGNRFKVEHYAVEPLNAVERGIVEQVGEAIRAVSRSGPKFAFAA 122
 QY 138 FSAHLNSAIKPIKPLP-----HLIKAQRE-KRFRYSY----- 169
 Db 123 GSAVITKLI-----PMPVGLDEQDLAEQIEIETATNIPYIEEVSLDFEVLGVPNNTEMV 178
 QY 170 -LLLR-----ELRK-AIKLVFEGKVTAKVKDIEAVPVWVAVNTAVMLIGRLPLPNPKK 220
 Db 179 QVLLAASRSNVELRQSALEL---GGLTAKVIDVEA---LAVENAFSLIAQELS VGSNA 231
 QY 221 VASWIEDKNILLYGTIDIEFGYRDIAGRSVGLLEVIDELNSELCPSELKHSGRELY 280
 Db 232 LVALI-----DIGATWSTLNLV-----HSGRSLYTR 257
 QY 281 TSSWADKSLRIWREDEGNARLNMLYMRGE 310
 Db 258 EQLFGKQL-----TDEVHRYGMYEEAGQ 283

RESULT 9

B64501

alpha-amylase (EC 3.2.1.1) - Methanococcus jannaschii

C:Species: Methanococcus jannaschii

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000

C:Accession: B64501

R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak

; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,

rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.

Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese

A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc

A:Reference number: A64300; MUID:96337999

A:Accession: B64501

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-467 <BUL>

A:Cross-references: GB:U67601; GB:L77117; NID:g2826439; PIDN:AAB99631.1; PID:g1592212

C:Genetics:

A:Map position: FOR1586396-1587799

C:Function:

A:Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds

Db 266 ALTSPKFAKARNWELGTSIDVDAVGSNIRIDSRGPEVMRVVPRNLNEDINEEWISDK 325
QY 229 NILLY-----GTDIEF--IGYRDIAGRMSVEGLLEVIDELNSLCPSEL-- 270
Db 326 TRFFYDGLKQRLNDPMIRGADGRFOAVSWRD-----ALAIYVAEYWHQIKPEIVG 376
QY 271 ---KHSGRELYLRTSSWADK--SLRTWREDEG---NARLNMLYNMGEALFAIENSDA 320
Db 377 VAGKLSAESAEMWALKDLLNKGNNIFCEGNGMHPNADLRSGYIMNTSISGL--EKADA 433

RESULT 12

T08162

amylopullulanase (EC 3.2.1.-) - Pyrococcus furiosus

C:Species: Pyrococcus furiosus

C:Date: 21-May-1999 #sequence_revision 21-May-1999 #text_change 08-Oct-1999

C:Accession: T08162

R:Dong, G.; Vieille, C.; Zeikus, J.G.

Appl. Environ. Microbiol. 63, 3577-3584, 1997

A:Title: Cloning, sequencing, and expression of the gene encoding amylopullulanase from
A:Reference number: Z16389; MUID:97438521

A:Accession: T08162

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-853 <DON>

A:Cross-references: EMBL:AF016588; NID:g2435436; PIDN:AB71229.1; PID:g2435438

C:Gene: apu

C:Genetics:

C:Function:

A:Description: hydrolyzes alpha-1,6 and alpha-1,4 linkages in starch-related polysacchar
C:Keywords: glycosidase; hydrolase; polysaccharide degradation

Query Match 5.7%; Score 102; DB 2; Length 853;
Best Local Similarity 21.2%; Pred. No. 8.3;
Matches 90; Conservative 60; Mismatches 155; Indels 120; Gaps 22;

QY 11 QYAEIPKSEPKV-----IEKAYI-----PVIETLIKEEPFGLNIT 45
Db 172 KYANLPLEEQKAVATNEFTQDYIDLAVLFNLAWIDYNIISTPELKALYDKVDEG--- 227
QY 46 GYTLKFLPK-----DILLVKGGIASDLIEIGTSYTAILPLPL-- 84
Db 228 GYTREDLKVLYHOMLLNNTFKBEKINLLNGN-----VEVTVPYAH--PIGILN 280
QY 85 -----SRVEAOQVRORVKEELFE-----VSPKGFWLPE--LADPIIPAILKDNNGEYL 130
Db 281 DFGWSEDFDAHVK--AHELYKYLGGVATPRGGWAASALNDKTL--EILAEQWQW 336
QY 131 FADAML-----FSAHLNSAIKPIKPLPHLIKAQREKRRFYISYL----- 170
Db 337 MTDQVLERMGIPYSIENYRPPWAEFNG--KKIYLFPRNHLSDRVGFYSGMNQYEV 394
QY 171 --LLRELKAIKLVFEGKVTIKVKDIEAVPVWAVNTAVMLIGRLPLMPKKVASWIECK 228
Db 395 EDFINELLKTKQYNDGL--VYVITLDGENPW-----EHYPYDGKLFLETLYKRLSELQEA 449
QY 229 NILLYGTDIEFIG--YRDIAGRMSVEGLLEVIDELNSLCPSEL--ELKHSGRELY----- 278
Db 450 GLIRLTPTPTVYIQLYGDKANKLPQ--MWERLDTTEERVEALKVANSGLGELYDLAGVTEE 508
QY 279 --LRTSSWADKSLRIW--REDEGNARLNMLYNMGEALFAIENSADGWPLPERRLDAFRA 335
Db 509 MQWPESWDGTLTSTWIGEPQENYATWYLYLARRTLMEKNKDKMSASWEKAYEYL--LRA 566
QY 336 IYNDW 340
Db 567 EASDW 571

RESULT 13

T48464

hypotheical protein TIE3.40 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48464

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, April 2000

A:Reference number: 224491

A:Accession: T48464

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-705 <BEV>

A:Cross-references: EMBL:AL162972

A:Experimental source: cultivar Columbia; BAC clone TIE3

C:Genetics:

A:Map position: 5

A:Introns: 59/1; 286/1; 350/1; 493/3

A:Note: TIE3.40

Query Match 5.7%; Score 101.5; DB 2; Length 705;

Best Local Similarity 21.2%; Pred. No. 6.9;

Matches 81; Conservative 59; Mismatches 141; Indels 101; Gaps 20;

QY 1 LRALVFHGNLQYAE-IPKSEPKVIE---KAYIPVETLIKEEPFGLNITGYTLKFLPKD 55
Db 143 LTVVAVSGNMEIAEALVAKNPKELEIPGINGQIPVV---VAVENTQEMARYLYTRTPVQ 199
QY 56 IILVKGGIASDLI-----EIGTSYTAILPLPLSRVEAQ-----VQRDRVKEELFEVS 104
Db 200 VLLAEDGYHTGLFLNLAIFYRMGLKGFLGQATHIFGGFDLYLFFFTOLD--IALDLFNMS 258
QY 105 PKGFWLPPELADPIIPAILKDNNGEYLFADFAFLFSAHLNSAIK--PIKPLP--HLIKAQREK 162
Db 259 RRLAVIKHLOIESIPII-----VLAKSPDLFP---GIQVKLPTLPKPSHANKDKSK 307
QY 163 RFRYIS-----YLLRELKAIKLVFEGKVTIKVKDIEAVPVWAVNTAVMLIGRLPLM 216
Db 308 FFRHIVKYKSIYIPLKVKRKSDFLFPD---TLMRKLLKGLSKWTGIDE---VYRLKVM 360
QY 217 N--PKKVASWIEDKNILL-----YGTDIEFIGYRDIAGRMSVEGLE 256
Db 361 HLQAKLLIGISEETLGLKERSETVDALLFAVRYG-----NVDFLVE 405
QY 257 VIDELNSLCPSELKHSGRELYLRTSSWADKSLRIWREDEGNARLNMLYNM--RGLAPL 314
Db 406 MTRN--NSELWSTRTSSSTLFLLAVERQEV-----FNLLYGLDDRKYLLLA 453
QY 315 AENSADAR-----GWPLPERRL 330
Db 454 DKDSGNGVHLIAGFPSPPSKL 475

RESULT 14

F81686

DNA-directed RNA polymerase, beta' chain TC0588 [imported] - Chlamydia muridarum (str

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000

C:Accession: F81686

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke

, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

A:Reference number: A81500; MUID:20150255

A:Accession: F81686

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1396 <TET>

A:Cross-references: GB:AE002328; GB:AE002160; NID:g7190627; PIDN:AAF39420.1; PID:g719

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

C:Gene: TC0588

C:Superfamily: Escherichia coli DNA-directed RNA polymerase beta' chain

Query Match 5.7%; Score 101.5; DB 2; Length 1396;
Best Local Similarity 22.3%; Pred. No. 18;
Matches 84; Conservative 62; Mismatches 142; Indels 89; Gaps 18;

QY 15 IPKSEPKVIEKAYIVETLKEEPPGLN--ITGVTFLKPLKDIILVKGGIASDLIEIIG 72
DB 404 IQGCAPEVND-----VLEELIKGHPVLLNRAPTLHRLGTAQAEFVLEIEGK-AIRVHPLVC 457
QY 73 TSSTA-----LPLPLSRVQAQVORDRKEELFEVSPKGFWLPDLADPI-IPAILKDN 125
DB 458 AAFNADFDGQMAVHVPLS-IEAQL-----AKVLMADPNIFLPSGKPVATPSKDMTL 511
QY 126 GYELFADEAMLFSAHLSAIPKPLPLIKA-----OREKRFYISYL 170
DB 512 GIYLLMADPTYPEEH-GGKTVKDEVEVLNALNAGGFILKDEICGSRDETGRGI--- 567
QY 171 LLRELKAIKLVPEGRVTKLVKDIEAVPVVAVNTAV-----MLIGRLPLMN 217
DB 568 ---HIHEAIKVRIDGQI-----IETTPGRVFETIVPKELGFQNYMPSKRISELILQC 618
QY 218 PKVASWIEDKNLLYGTIEFGYRDIAGRMSVEGLLEV-IDELNSELCPSELKHSGRE 276
DB 619 YKKVGL-----EATVRFDLDELKELGFVQSTKAAISMGLKDVRIPEIKKEI---LKDAYDK 670
QY 277 LYLTSSWAD-----KSLRIWREDEGNARLNMLY-----NMRGELAFLAENS 319
DB 671 VAVVKQYEDGIITDGERHSKTIISWTE-VSDLLSNALYAEIKQTNKSHNPLFLMIDSG 729
QY 320 ARGWPLPERRLDARAI 336
DB 730 ARGNKSQKOLGALRGL 746

RESULT 15

F75006
hypothetical protein PAB1309 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: F75006
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: F75006
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <KAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:g5458960; PIDN:CAB50492.1; PID:g545900
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1309

Query Match 5.7%; Score 101; DB 2; Length 394;
Best Local Similarity 21.5%; Pred. No. 3.3;
Matches 53; Conservative 43; Mismatches 66; Indels 84; Gaps 13;

QY 7 HGNLQYAEIPKSEPKVIEKAYIVETLKEEPPGLNITGTLKPLKDIILVKGGIASD 66
DB 72 HGN-----IVLALSPIIGLI-----GLGI-GYFLY-----TGGWASG 104
QY 67 LIETIGTSYTAILPLPLSRVQAQVQDRVKEELFEVSPKGFWLPDLADPIIPAILKDN 126
DB 105 DVIILG-AYSALLPYVP-----DSAKY-----PPYSYILPM-----NA 137
QY 127 YEYLFADAMLFSAHLSAIPKPLPLIKAQREKRFYISYLLRELKAIKLVFEKG 186
DB 138 FTILFNSLLIFPLILYVS-----VGLAVKGKIRELVGVFREG- 176
QY 187 VTLKVDIEAVPVVAVNTAVMLIGRLPL-MNPKKVASWIEDKNLLYGTIEFGYRDI 245
DB 177 ----IRNVVEITLW--INFGAVLLAFVSLHMSIPKTIISWLTFTALILFSRKLVG--DV 228

QY 246 AGRMSV 251
DB 229 LGVLSI 234

Search completed: April 4, 2002, 14:41:20
Job time: 184 sec

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	151	8.5	685	1	AMV1_DICTH	P09961 dictyoglomu
2	126.5	7.1	633	1	AMV1_PYROH	O57932 pyrococcus
3	120.5	6.8	655	1	AMV1_PYRAB	P9V298 pyrococcus
4	116.5	6.6	648	1	AMV1_PYRFU	Q49067 pyrococcus
5	111	6.2	4725	1	DYHC_DICTDI	P34036 dictyosteli
6	106.5	6.0	467	1	AMV1_METJA	Q59006 methanococ
7	104.5	5.9	1250	1	BXE_GLOBU	P30995 clostridum
8	102	5.7	738	1	NUAM_SOLITU	Q43644 solanum tub
9	101.5	5.7	1396	1	RPOC_CHLMU	Q9pk79 chlamydia m
10	100.5	5.7	653	1	MALQ_PYROK	O32450 pyrococcus
11	99.5	5.6	630	1	Y235_METJA	Q60291 methanococ
12	97.5	5.5	659	1	MALQ_THELI	O32462 thermococcu
13	97.5	5.5	1396	1	RPOC_CHLTR	O84316 chlamydia t
14	97	5.5	924	1	DPOL_BPSP1	P30314 bacteriopho
15	96.5	5.4	1367	1	DPO3_THEMA	Q9zhf6 thermotoga
16	96	5.4	1210	1	YEHL_ECOLI	P33346 escherichia
17	96	5.4	1242	1	MSH6_YEAST	O03834 saccharomyc
18	95.5	5.4	471	1	UDPG_PYRYP	O64459 pyrus pyrif
19	94	5.3	467	1	UDPG_MUSAC	Q9sdx3 musa acumin
20	93	5.2	473	1	UDPG_HORVU	Q43772 hordeum vul
21	93	5.2	681	1	DP3X_MYCPV	P75177 mycoplasma
22	92	5.2	571	1	ILV1_BUCAL	P57321 buchnera ap
23	91	5.1	598	1	Y288_THEMA	Q9wyc4 thermotoga
24	90.5	5.1	781	1	Y5B_YEAST	P53316 saccharomyc
25	90	5.1	470	1	PPOX_BACSU	P32397 bacillus su
26	89.5	5.0	387	1	HEMZ_SYNY3	P54225 synechocyst
27	89.5	5.0	446	1	R554_BACSU	P37105 bacillus su
28	89.5	5.0	1078	1	RPOB_SINAL	P46818 sinapis alb
29	88.5	5.0	447	1	SR54_MYCWI	Q01442 mycoplasma
30	88	4.9	257	1	VBR1_TMOV	Q06661 tomato mott
31	88	4.9	572	1	PT1_STAAR	P51183 staphylococ
32	88	4.9	592	1	STB3_HUMAN	Q00186 homo sapien
33	87.5	4.9	476	1	HDPG_SOLITU	P19595 solanum tub

Query Match 7.1%; Score 126.5; DB 1; Length 633;
 Best Local Similarity 22.3%; Pred. No. 0.026;
 Matches 87; Conservative 59; Mismatches 121; Indels 123; Gaps 24;

QY 56 ---IILVKGASDLIEICTSYTAILPLLSRVEAQVDRVKE-----ELFEV 103
 Db KLIKMAERGQIEF---VSGGFYPIIP-----DKDKVQIKKLNKYIYDKFG 115

QY 104 SPKGFWLPE-LADPIIPAILKDNQYEVLFADAMLSAHLNSATKPKPLPHLIKAQREK 162
 Db TPKGWLAERWEPHLVYIAEAGIEVYVDDAHFFSVGL-----K 156

QY 163 RFRYSIYLLBELKAIKL-VFEGKVTLKVDIEAVPVAVNTAVMLIGRLPLMPKVV 221
 Db EEDIFGYLMEE---QGYKLVF---PISMKLRYL-----IPFADPEET 194

QY 222 ASWI-----EDKN---ILLYGTDTIEFGYRDIAGRMSVEGLE-----VIDELNS 263
 Db ITYLDKFASEDKSKIALLFDDGKFGWLPDPTYRTVVEGWLTFVSKIKENFLVTPVNL 254

QY 264 ELCPELKHSGRELYLRTSSW-----ADKSLR-----290
 Db 255 YTYMORVKPKGR-IYLPASRYEMWVLPPEAQKEELVEKLTKENLWDKFSPPYVKG 313

QY 291 IWRB-----DEGN-ARLNMLYNNRGAFLAENSADRGWPLPERRL-DAFRAIYND--WR 341
 Db 314 FWRNLAKYDESNMOKKMLYVWK-----KYQDSPNEEVKEKAMEEVFQGOANDAYWH 366

QY 342 G 342
 Db 367 G 367

RESULT 2
 ID AMYA_PYRHO STANDARD; PRT; 633 AA.
 AC 057932;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALPHA-AMYLASE (EC 3.2.1.1).
 GN AMYA OR PH0193.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
 RL DNA RES. 5:55-76(1998).
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
 CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AP000001; BAA29262.1; -
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
 SQ SEQUENCE 633 AA; 75017 MW; 9DCBCB2FE191501 CRC64;

Query Match 7.1%; Score 126.5; DB 1; Length 633;
 Best Local Similarity 22.3%; Pred. No. 0.026;
 Matches 87; Conservative 59; Mismatches 121; Indels 123; Gaps 24;

QY 21 KVIEKAYIPVETLKEEP---FGLNITGYTLKFL---PKDILILVKGASDLIE-IIG 72
 Db 25 RAYEKAYRPFLETL-EEYDNMKVAVHISGLVFWLERNRPEYIDLKSLIKGQVELVVA 83

QY 73 TSYTAILPLP-LSRVEAQVDRVKEELFEVSPKGFWLPE-LADPIIPAILKDNQYEV 130
 Db 84 GYEPILVAIPEDRVE-QIKLSKGWARKMGYEARGLWTERVWPELVKTLREAGIEV 142

QY 131 FADAMLSAHLNS-----AIKPIKPLPHLIKAQREKFRYISVLLLE 174
 Db 143 ILDDHMSAGLSKEELFWPYTENGGAIVVFP-----DEKLRYL--IPFRP 189

QY 175 LKATKLVF-----EGVTLKVDIEAVPVAVNTAVMLIGRLPLMPKVKASWI--- 225
 Db 190 VNETLEYLHSLADEDESKVAVFDDGKFGAWPGTHLVYERG-----WLKEF 237

QY 226 -----EDK-NILLYGTDTIEFG-----YRDIAG--RMSVGL-----LEV 257
 Db 238 FDRISDDKINMLIS---EYLSKPRGLVLYPLTIASFENSEWSLSPARQAKLFEFFIK 294

QY 258 IDELSELCPSELKSHGRELRYLTSSWADKSLRIWREDEGN---ARLNMLYNNRGAFL 314
 Db 295 LKELN-----LFEKYRIFVRGINKN---FLYKYPEGNYMHRMLMLSKL-----L 337

QY 315 AENSARGWPLPERRLDFAIYND--WRG 342
 Db 338 RNN-----PTARIFVLVLAQCNDAYWHG 359

RESULT 3
 ID AMYA_PYRAB STANDARD; PRT; 655 AA.
 AC 09V298;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ALPHA-AMYLASE (EC 3.2.1.1).
 GN AMYA OR PA0118.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORSAY;
 RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
 structure and evolution.";
 RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
 CC -!- PATHWAY: POLYSACCHARIDE DEGRADATION.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ248283; CAB49100.1; -
 KW Hydrolase; Glycosidase; Carbohydrate metabolism; Complete proteome.
 SQ SEQUENCE 655 AA; 77296 MW; 7F6F920B1A00EECE CRC64;

Query Match 6.8%; Score 120.5; DB 1; Length 655;
 Best Local Similarity 20.1%; Pred. No. 0.079;
 Matches 79; Conservative 62; Mismatches 117; Indels 135; Gaps 20;

QY 73 TSY-TAILPLPLSRVQAQVDRVKEELFEVSPKGFWLPELADPIIPAILKONGYEYLF 131
 Db 373 EGYPTTELL-----ATRIFFPGRSLSLSLAR-----DLG-VVP-----GY-FLF 410
 QY 132 ADEAML-----FSAHLSAIPKIPKPLPHLIKAQREKFRFYISYLL-----RELKKA 178
 Db 411 SDVAILVSGNSDRILQYIDLLTEQLKNLPK-----FVLIIPYQARSLSKEM 458
 QY 179 IKLVPEGVKTLVKVDIEAVPVAVTAVMLIGRLPLMNPVKVASWTDKNI-----L 231
 Db 459 IK-VFKEDYGKIV-----FLIADNDEGL-----KNVKTAKMDGL 492
 QY 232 LYGTDIEPIGVDRVIAGRMSVGLLEVIDELNSELCPSELKSHGRELVLRTSSWADKSLRI 291
 Db 493 RENFEVFLIEKKDILGYIKPDPFLALDELLRECKDKYEEVLODEIKSIIDEIKFIGA 552
 QY 292 WREDEGNARLNMLNMRGELAFLAENSAR-----GWPLPERRLDAPRAI 336
 Db 553 CKNEEGKNEKDLHLKSLKFLTVLKNVDEEFITLW-----DSGRSI 595

RESULT 12
 MALQ_THELI STANDARD; PRT; 659 AA.
 AC 032462;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE 4-ALPHA-GLUCANOTRANSFERASE (EC 2.4.1.25) (AMYLOMAL/TASE)
 DE (DISPROPORTIONATING ENZYME) (D-ENZYME).
 GN JGT.
 OS Thermococcus litoralis.
 OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
 OX NCBI_TaxID=2265;
 RN [1]
 RP SEQUENCE FROM N.A., SEQUENCE OF 1-19 AND 427-437, AND CHARACTERIZATION.
 RC STRAIN=DSM 5473;
 RC MEDLINE=97454300; PubMed=9310375;
 RA Jeon B.-S., Taguchi H., Sakai H., Ohshima T., Wakagi T., Matsuzawa H.;
 RT "4-alpha-glucanotransferase from the hyperthermophilic archaeon
 Thermococcus litoralis. Enzyme purification and characterization, and
 RT gene cloning, sequencing and expression in Escherichia coli.";
 RL Eur. J. Biochem. 248:171-178(1997).
 CC -!- FUNCTION: CATALYZES THE TRANSGLYCOSYLATION OF
 MALTOOLIGOSACCHARIDES, YIELDING MALTOOLIGOSACCHARIDES OF VARIOUS
 LENGTHS AND GLUCOSE.
 CC -!- CATALYTIC ACTIVITY: TRANSFERS A SEGMENT OF A 1,4-ALPHA-D-GLUCAN TO
 A NEW 4-POSITION IN AN ACCEPTOR, WHICH MAY BE GLUCOSE OR 1,4-
 ALPHA-D-GLUCAN.
 CC -!- ENZYME REGULATION: INHIBITED BY P-CHLOROMERCURIBENZOIC ACID,
 MONOIODOACETIC ACID, MERCURY AND NICKEL IONS.
 CC -!- MISCELLANEOUS: OPTIMAL ACTIVITY IS FOUND AT 90 DEGREES CELSIUS.
 CC -!- SIMILARITY: BELONGS TO FAMILY 57 OF GLYCOSYL HYDROLASES.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; D88253; BAA22063.1; -
 KW Transferase; Glycosyltransferase; Carbohydrate metabolism.
 FT ACT_SITE 352 352 PROBABLE.
 SQ SEQUENCE 659 AA; 77885 MW; F789AFF9BF8281AC CRC64;

Query Match 5.5%; Score 97.5; DB 1; Length 659;
 Best Local Similarity 18.6%; Pred. No. 4.9;
 Matches 70; Conservative 69; Mismatches 134; Indels 103; Gaps 17;

QY 25 KAVIPVIETL--IKKEPFGCLNITGYTKFL-----PKDILVKGGIASDLIEIGTS-YTA 77
 Db 29 RSYRPWEILEEFEPKMNKVVHFGSPGLEWEEKPDYDLRLSLKRGQLEIVWAGFVEP 88
 QY 78 ILPLPLSRVQAQVDRVKEELFEV--SPKGFWLPE-LADPIIPAILKONGYEYLF 134
 Db 89 VLAaip--KEDRLVQIEMKDYARKLYGDAKGVWLTERVWQPELVKSLREAGIEYVWVDD 146
 QY 135 AMLFSAHLNS-----AIKPIKPLPHLIKAQREKFRFYISYLLRELKKA 178
 Db 147 YHFMASGLSKEELFWPYTYTDEGGEVITFPI-----DEKRLV--IPFRPVKKT 193
 QY 179 IKLVPE-----GKVTLVKVDIEAVPVAVTAVMLIGRLPLMNPVKVASWTDKNI 232
 Db 194 IEYLESLSDDPSKAVAFHDDGKFKGVWPG-----TYEWVYKGVWL 235
 QY 233 YGTD-----IEFGYRDIAGRMSVGLLEV-----IDELNSELCP 270
 Db 236 EFFDAITSNEKINLMYSEYLSKFTPRGLVYLPASIFEMSEWSLPKQAKLFVEFVEQL 295
 QY 271 KHSGR-----ELYRTSSWADKSLRWREDEGNARLNMLNMRGELAFLAENSAR 326
 Db 296 KEEGPEKYRVFVGIGWKNFFKYPESFMHMKRLMVSKA-----VADNPEARKYILK 349
 QY 327 ERRDLAFRAIYNDWRG 342
 Db 350 AQONDAY-----WHG 359

RESULT 13
 RPOC_CHLTR STANDARD; PRT; 1396 AA.
 ID RPOC_CHLTR
 AC 084316;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE DNA-DIRECTED RNA POLYMERASE BETA' CHAIN (EC 2.7.7.6) (TRANSCRIPTASE
 BETA' CHAIN) (RNA POLYMERASE BETA' SUBUNIT).
 DE RPOC OR CT314.
 GN Chlamydia trachomatis.
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OC Chlamydia trachomatis.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RC MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger R.L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
 RNA(N).
 CC -!- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
 ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
 BETA' CHAIN.
 CC -!- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; AE001304; AAC67907.1; -
 DR InterPro: IPR000722; RNA_pol_A.
 DR Pfam: PF00623; RNA_pol_A; 1.

"The hyperthermophilic bacterium *Thermotoga maritima* has two different classes of family C DNA polymerases: evolutionary implications."; *Nucleic Acids Res.* 26:5300-5309(1998).

[2] SEQUENCE FROM N.A.
STRAIN-MSB8 / DSM 3109;
MEDLINE=99287316; PubMed=10360571;
NELSON K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Hart D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Saizberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
"Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of *Thermotoga maritima*."

Nature 399:323-329(1999).
 FUNCTION: REQUIRED FOR REPLICATIVE DNA SYNTHESIS. THIS DNA POLYMERASE ALSO EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
 CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE + DNA(N).
 SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
 SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-C FAMILY. POLC SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@sib-sib.ch).

EMBL; AF065313; AAC80438.1; -
EMBL; AE001732; AAD35661.1; -
TIGR; TM0576; -

InterPro: IPR002562; 3_5_exonuclease.
InterPro: IPR000520; Exonuclease.
InterPro: IPR003141; p11p_N.

InterPfo: IPR002309; tRNA-synt_2.
pfam: PF00929; Exonuclease; 1.
pfam: PF02231; PHP_N; 1.
pfam: PF01336; tRNA_anti; 1.

SMART; SM00474; 35EX0C; 1.
SMART; SM00479; EX0III; 1.
SMART; SM00481; POLIIIAC; 1.
Transferase; DNA-directed DNA polymerase; DNA replication; Hydrolase;

Nuclease; Exonuclease; Complete proteome.
DOMAIN 358 521 EXONUCLEASE.
SEQUENCE 1367 AA; 155362 MW; EE5916FA70591F84 CRC64;

every Match	5.4%	Score 96.5;	DB 1;	Length 1367;
1st Local Similarity	22.7%	Pred. No. 16;		
atches 96: Conservative	58:	Mismatches 154:	Indels 115:	Gaps 20:

3 ALVFHGNLQ---YAEIPKS---EPKVEIKAYI-----PVIETTLIKERPPGINITGYTLK 50
304 ALTDHCNVCIAIDVEVDAAKFAKCIKFCIEFVILVSNVDPVIRNI SDNSTEC-DATEVID 362

[illegible]

```

363 FETTGDPQVDIEIIEIGAVKIQGGQIVDEYHTLTKPSREISRKSKSEITGITQEMLENKRS 422
100 LFEVSPKGFWLPELADPIIPAILKONGVEYL-----FADEAMLF----- 138
: || : : | | | | : | : |

```

423 IEEVLE--FLOFLEDSIIVAHNANFYRFLWIKKVGMDWERPYDITLALAKSLKL 480

139 -SAHLNSAIKP--IKPLPHLIKAQREKFRYISYLLLRKAIKLVFEGKVTLVKDIE 195

```

481 RSYSLDSVVEKLGCGFRH-HRALDARVTAQVLFREVMKKIGITKLEWE-KLRDTI 538
196 AVPVVVVAVNTAVMLGRUPLMPPKVAS-----WIEDKNLLYGTDIE 238
      : :::: : | | | : |

```

— — — — —

100

Db	539	DYATKPFHCTILVQNKKGKGLKNLYKLVSDSYKYFYFVGPVRIKLSELIEENREGLLVGSAC-	597
QY	239	FIGYRDITAGMSVEGL-----LEVIDELNSELCPSELKHSGRELYL	279
Db	598	---ISGLGRAALEGASDSELEETIAKFYDIEVWPLDVIAEEDLDRELRKEVVKLY-	653
QY	280	RTSSWADKSLRIWREDEGNARLNWLYNMGETAFLAENSARG-----WPLPERLDAPR	334
Db	654	-----RIAK-----KLNKFVVMTGDVHFL- DPEDARGAALLAPOGNRNFENQP	696
QY	335	AIY	337
Db	697	AIY	699

Search completed: April 4, 2002, 14:44:54

Job time: 248 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:40:16 ; Search time 28.84 Seconds
(without alignments)
1754.862 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 1778
Sequence: 1 LRALVHGNLQYAEIPKSEP.....ERRLDAPRAIYNDWNGNEP 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organella.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.rodent.*
- 12: sp.virus.*
- 13: sp.vertibrate.*
- 14: sp.unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1268	71.3	364	1 Q9HHB5	Q9hhb5 pyrococcus
2	1229	69.1	364	1 O58106	O58106 pyrococcus
3	110	6.2	1362	1 Q9V294	Q9v294 pyrococcus
4	106.5	6.0	392	2 Q9RGD3	Q9rgd3 xyella fas
5	106	6.0	458	10 Q9MBD0	Q9mbd0 pyrus pyrif
6	104.5	5.9	1255	2 Q9FAR6	Q9far6 clostridium
7	102	5.7	853	1 Q30772	Q30772 pyrococcus
8	101.5	5.7	705	10 Q9LZ29	Q9l229 arabidopsis
9	101	5.7	394	1 Q9UYC0	Q9uyco pyrococcus
10	100.5	5.7	653	1 Q32450	Q32450 pyrococcus
11	100.5	5.7	690	2 Q9PKB1	Q9pkb1 chlamydia m
12	100	5.6	336	1 Q58818	O58818 pyrococcus
13	99	5.6	391	1 Q58793	Q58793 methanococ
14	99	5.6	608	10 Q9AVF4	Q9avf4 anaranthus
15	99	5.6	742	5 Q17474	Q17474 hydra oliga
16	98.5	5.5	1518	13 Q9IAR8	Q9iar8 salmo salar
17	98	5.5	1430	2 Q9PA87	Q9pa87 xyella fas
18	97.5	5.5	564	2 Q99WX0	Q99wx0 staphylococ
19	97.5	5.5	659	1 Q32462	Q32462 thermococcus

20 97.5 5.5 1473 3 Q14167
21 97.5 5.5 1503 13 Q73677
22 97 5.5 480 1 Q9YDN6
23 97 5.5 1511 13 Q9I8E2
24 96.5 5.4 819 2 Q51074
25 95.5 5.4 588 3 Q13693
26 95.5 5.4 684 2 Q9M2S4
27 95.5 5.4 1285 2 Q9MXU3
28 95 5.3 445 2 Q9I012
29 95 5.3 1337 1 Q9Y8I8
30 94.5 5.3 317 1 Q58067
31 94 5.3 471 10 Q9MA46
32 94 5.3 1251 2 Q9K395
33 93.5 5.3 235 10 Q9M276
34 93.5 5.3 641 1 Q58428
35 93 5.2 497 5 Q27420
36 93 5.2 923 2 Q9KD04
37 93 5.2 948 2 Q9H2E4
38 92.5 5.2 4589 5 Q76506
39 91.5 5.1 364 2 Q31292
40 91.5 5.1 520 2 Q9PMZ7
41 91.5 5.1 529 2 P74630
42 91.5 5.1 794 2 Q9X2I2
43 91 5.1 510 2 Q26022
44 91 5.1 512 4 Q9H5U5
45 91 5.1 825 5 Q16769

ALIGNMENTS

RESULT 1
Q9HHB5 PRELIMINARY; PRT: 364 AA.
ID Q9HHB5
AC Q9HHB5;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE ALPHA-GALACTOSIDASE.
GN GALA.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3638;
RA Verhees C.H.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF195244; AAG28455.1; -
SQ SEQUENCE 364 AA; 41545 MW; 3B8E0AF5BDCCF2A5 CRC64;

Query Match 71.3%; Score 1268; DB 1; Length 364;
Best Local Similarity 75.2%; Pred. No. 2.2e-81;
Matches 273; Conservative 38; Mismatches 34; Indels 18; Gaps 18;
QY 1 LRALVHGNLQYAEIPKSE-PKVIKAYIPVETLKEE-PGLNITGYTLKFLPKDII- 57
DB 1 MRLVHGNLQYAEIPKSEIPKVIKAYIPVETLKEE-PGLNITGYTLKFLPKDIIA 60
QY 58 LVKGGIADLIETIGTSTY-AILPLPLSRVDAQVQRD-VKEELFEVSPKGFWLPELA- 114
DB 61 LIKEGIESGLIELGTSYTHAILPLPLSRVDAQVQRD-VKEELFEVSPKGFWLPELAY 120
QY 115 DPTIIPAILKDNQYGYLFAD-EAMLSAHLNSAIKPKPL-PHLIAQREKRPRTSYLL- 171
DB 121 DPTIIPAILKDNQYGYLFAD-EAMLSAHLNSAIKPKPL-PHLIAQREKRGVLVLYLLG 180
QY 172 LRELKAIKLVPEGKWTLK-VKDIEAVPVVAVNTAVML-IGRLPLMPPKVKVSDIKN 229
DB 181 LRELKAIKLVPEGKWTLEAVKEIEAIPVWSINTAVMLGAGRFPLMPPKVKVAKWKBKD 240
QY 230 -ILLYGTDIEFGYRDYAG-RMSVEGLLEVIDELNSEL-CPSELKHSGRLEYLRTSSWA- 285

Db	241	EILLYGTDFELGYRDIAGYKIIISNIINLEEGELGUPRKIKHSEKKLYLTSSWAP	300
Qy	286	DKSLRIWREDEGNARLNMLYN-MRGELAFIAENSARDGW-PLPERLDAFRAIYNDRG-	342
Db	301	DKSLRIWTEDEGNARLNMLTSDYMGELAFIAENSARDGWEPLPERLDAFKAIYTHWSE	360
Qy	343	NGE 345	
Db	361	NGK 363	
RESULT	2		
O58106		PRELIMINARY;	
ID	O58106	PRT;	364 AA.
AC	O58106;		
DT	01-AUG-1998	(TREMBLrel. 07, Created)	
DT	01-AUG-1998	(TREMBLrel. 07, Last sequence update)	
DT	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
DE	HYPOTHETICAL 41.8 KDA PROTEIN PH0368.		
GN	PH0368.		
OS	Pyrococcus horikoshii.		
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.		
OX	NCBI_Taxid=53953;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=OT3.		
RC	MEDLINE=98344137; PubMed=9679194;		
RA	Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,		
RA	Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,		
RA	Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,		
RA	Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,		
RA	Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,		
RA	Masuchi Y., Shizuya H., Kikuchi H.;		
RT	"Complete sequence and gene organization of the genome of a hyper-		
RT	thermophilic archaeobacterium, Pyrococcus horikoshii OT3."		
RL	DNA Res.5:55-76(1998).		
DR	EMBL: AF000002; BAA29442.1; -.		
DR	Hypothetical protein; Complete proteome.		
SW	SEQUENCE 364 AA; 41755 MW; 7B4B36AB4A975BAD		CRC64;

[illegible]

Q9MBO0	PRELIMINARY;	PRT;	458 AA.
Q9MBD0			
01-OCT-2000	(TREMBLrel. 15, Created)		
01-OCT-2000	(TREMBLrel. 15, Last sequence update)		
01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
UDP-GLUCOSE PYROPHOSPHORYLASE			
Pyrus pyrifolia (Japanese pear) (Pyrus serotina).			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatids I; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
eurosid 1; Rosales; Rosaceae; Pyrus.			
NCBI_TaxID=3767;			
SEQUENCE FROM N.A.			
STRAIN-CV, NIJISSEIKI;			
Norioka S., Kiyozumi D., Norioka N.;			
"Molecular cloning and nucleotide sequencing of a gene encoding UDP-			
glucose pyrophosphorylase of Japanese pear (Pyrus pyrifolia Nakai).";			
Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.			
EMBL; AB027617; BAA96250.1; -			
InterPro; IPR002618; UDPGP.			
Pfam; PF01704; UDPGP; 1.			
SEQUENCE 458 AA; 50771 MW; 7DA5BFCAC7F45850 CRC64;			
Query Match 6.0%; Score 106; DB 10; Length 458;			
Best Local Similarity 24.7%; Pred. No. 15;			
Matches 71; Conservative 37; Mismatches 91; Indels 88; Gaps 17;			
4 LVFHCNLOYAEIPKSEPKVIEKAVIPVETLKEEPPG--LNITGYTLKFLPKDILVKG 61			
55 VPHAGLEPA--PKDTAEI--KALLDKLVKLVKLGGLTGMCTG-----PKSVIEVRD 104			
62 GIA-SDL----IEIGTSYTAILPLPL-----SRVEAQV-----QRD 94			
105 GLTFLDLIVMGIHLNDKFGSSVPLLMSFNTHDDTKIVKYSKANVQIHTFNQSQYP 164			
95 RVKEELFVSP-----KGFWLPELADPIPA-----ILKNDGYEYLFADEAMLFSA 140			
165 RLVDVDDFTPLPSKQGTGKDGWYPGHGVPFSLKNSKLDLLLSQGREYVFSN----SD 220			
141 HUNSAIKPIKPLHPLIKAKOERKFRYSYL---LLRELRKAIKLVFEGKVTL----- 189			
221 NL-GAIVDLKILHLIKNKE---YIMEVTPKTLADVKGGILISYEGKVLLEIAQVPE 275			
190 ----KVKDIEAVPV-----WVAVNTAVMLIGRLPLM-----NPKKV 221			
276 EHVDFEKSIEKFKIPNTNNLWNLNAIKRLVDADALKMEIIPNPKV 322			
RESULT 6			
Q9FAR6	PRELIMINARY;	PRT;	1255 AA.
AC Q9FAR6;			
DT 01-MAR-2001	(TREMBLrel. 16, Created)		
DT 01-MAR-2001	(TREMBLrel. 16, Last sequence update)		
DT 01-JUN-2001	(TREMBLrel. 17, Last annotation update)		
DE TYPE E BOTULINUM TOXIN.			
GN BONT/E.			
OS Clostridium butyricum.			
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;			
OC Clostridium.			
OX NCBI_TaxID=1492;			
SEQUENCE FROM N.A.			
STRAIN-BL 6340/ATCC 43755/BL 5520/KZ 147;			
MEDLINE=20509829; PubMed=11055954;			
RA Wang X., Maegawa T., Karasawa T., Kozaki S., Tsukamoto K., Gyobu Y.,			
RA Yamakawa K., Oguma K., Sakaguchi Y., Nakamura S.;			
"Genetic Analysis of Type E Botulinum Toxin-Producing Clostridium			
butyricum Strains.";			
RT Appl. Environ. Microbiol. 66:4992-4997(2000).			
DR EMBL; AB039264; BAB12249.1; -			
DR InterPro; IPR000395; Bontoxilysin.			

```
DR InterPro: IPR000130; Zn_MTPetdse.
DR Pfam: PF01742; Peptidase_M27; 1.
DR PRINTS: PR00760; BONTOXILYSIN.
DR ProDom: PD001963; Bontoxilysin; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR SEQUENCE 1255 AA; 143918 MW; 1B557B9D85CD8E4D CRC64;

Query Match          5.9%; Score 104.5; DB 2; Length 1255;
Best Local Similarity 21.9%; Pred. No. 87;
Matches 89; Conservative 67; Mismatches 111; Gaps 22;

QY 10 LQYAEIPKPEPKVIEKAYIPVETLIKKEPFGNITGYTLKEL-----PKDIILVKGGA 64
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 526 LDAQKYPEGNNV---NLTSISDITALLEQ---KIYTFSSFEINNVKPVQAALFVGWIQ 580
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 SLDIETIG-----TSVTAILPLPLS-RVEAQVQDRVRKEELFEVSPKGFWL-- 110
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 581 QVLVDFTTEANQKSTVDKTDADISIVVYIGLALNIGNEAQGNFKDAL-ELLGAGILLEF 639
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 111 -PELADPIIPAILKONGYELFADEAMLSAHLNSAIKPKPLPHLIKAQREKFRYISY 169
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 640 EPEL---LIPTIL-----VFTIKFLGSS--DNKNVKAINKALKEDEKKEVYSF 687
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 170 LLRLRLK-----AKLVFEGK---VTLVKV-----DIEAV- 197
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 688 IVSNMWTKINTQFNKRKEQMYQALQNVNALKAIIESKYSNTLEEKNELNKVDIEQIE 747
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 198 -----PVWVAVN-----TAVMLIGRLPLMNPKKVASWIE-DKNILLYGTDIEFIGYDI 245
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 748 NELNQKVSITAMNIDRFLETSSISYLMKLINELVKNLREYDENVKTYLLDY-IKHGSI 806
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 246 AGRMSVEGLEVIDELNSELCPSELKHSGRELYLRTSSWADKSL-----RIWREDEGNA 299
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 807 LGSQQLNSMWIDFLNNSI-PFKL-----SSYTDKILISYFNKFFKRIKSSS 854
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 300 RLNNLY-----NMRGELAFLAENSADARGWPLPERLDAFRAIYND 339
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 855 VLNMRKNDKYVDTSYDSDNININGDVYKPTNKQF-----GIIND 896
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
O30772 ID O30772 PRELIMINARY; PRT; 853 AA.
AC O30772;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-AUG-1998 (TrEMBLrel. 07, Last annotation update)
DE AMYLOPOLLULANASE.
GN APU.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=2361;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 3638;
RX MEDLINE=97438521; PubMed=9293009;
RA Dong G., Vielle C., Zeikus J.G.;
RT "Cloning, sequencing, and expression of the gene encoding
RT amylopullulanase from Pyrococcus furiosus and biochemical
RT characterization of the recombinant enzyme.";
RL Appl. Environ. Microbiol. 63:3577-3584 (1997).
DR EMBL: AF016588; AAB71229.1; -.
SQ SEQUENCE 853 AA; 99035 MW; A25C2C21D2218165 CRC64;

Query Match          5.7%; Score 102; DB 1; Length 853;
Best Local Similarity 21.2%; Pred. No. 73;
Matches 90; Conservative 60; Mismatches 155; Indels 120; Gaps 22;

QY 11 QYAEIPKPEKV-----IEKAYI-----PVETLIKKEPFGNIT 45
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 172 KYANLPLEQKVAVTNEFTQDDYIDLAVLENLAWIDYNIISTPELKALYDKVDEG----- 227
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 46 GYTKELPK-----DIILVKGSIASDLIEITIGTSYTAILPLPL-- 84
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228 GYTREDLTKVLYHOMWLLNNTFKHEKINLLNGN-----VEVTVPYAH--PIGPILN 280
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 85 -----SRVEAQVQDRVRKEELFE-----VSPKGFWLPE--LADPIIPAILKONGYEL 130
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 281 DFGWSEDFDAHVKK---AHELYKKYGGGVATPRGGMAAESALNDKTL-ETLAENGWQW 336
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 131 FADAML-----FSAHLNSAIKPKPLPHLIKAQREKFRYISYL----- 170
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 337 MTDQWLMRGMGIPYSIENYRPWAEFNG--KKIYLFPRNHLDSDRVGFRYSGMNQVEAV 394
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 171 --LLRLRLKALKLVFEGKVTLVKVDIEAVPVWAVNTAVMLIGRLPLMNPKKVASWIEDK 228
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 395 EDFNELLKTKYNYDGLS-VYVITLDGENFW-----BHPYDGLFLETLYKRLSELQEA 449
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 229 NILLYGTDIEFIG-YRDIAGRMSVEGLEVIDELNSELCPSELKHSGRELY----- 278
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 450 GLITLTPTTEYIQLYGDKANKLTPO--WNERLDTTEERVEALKVANSGLGYLDLAGVTEE 508
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 279 --LRTSSWADKSLRIW-REDEGNARLNMLYNMRGELAFLAENSADARGWPLPERLDAFRA 335
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 509 MQWPSSWIDGTISTWIGEPQENYAWYLYLARRTLMENKDKMSASWEKAYEYL--LRA 566
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 336 IYNDW 340
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 567 EASDW 571

RESULT 8
Q9LZ29 ID Q9LZ29 PRELIMINARY; PRT; 705 AA.
AC Q9LZ29;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 79.5 KDA PROTEIN.
GN TIE3_40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Kouze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL: AL162972; CAB6010.1; -.
DR InterPro: IPR002110; ANK.
DR Pfam: PF00023; ank; 3.
DR PROSITE: PS0297; ANK_REGION; 1.
KW Hypothetical protein.
SQ SEQUENCE 705 AA; 79491 MW; 91C44E144E012CC7 CRC64;

Query Match          5.7%; Score 101.5; DB 10; Length 705;
Best Local Similarity 21.2%; Pred. No. 60;
Matches 81; Conservative 59; Mismatches 141; Indels 101; Gaps 20;

QY 1 LRALVFHGNLQVAE-IPKSEPKVIE-----KAYIPVETLIKKEPFGNITGYTKFLPKD 55
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 143 LTVAVSGNMEIAEALVAKNKLLEIPGNGQIPVW---VAVENTQMGMARYLTYRTPVQ 199
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 56 IILVKGSIASDLI-----EITIGTSYTAILPLPLSLRVEAQ-----VQDRVRKEELFEVS 104
   |:::| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Db 200 VLAEDGYHGTLLFLNALFYRMKGGFLGIGQATHIFGGFDLYLFFFIQD-LALDLFNMS 258
QY 105 PKGFWLPALADPIIPAILKONGYELFADEAMLSAHLNSAIK-PIKPLP-HLIKAQREK 162
Db 259 RELAVTKHLQIESPII-----VLASKPDLP---GIQVKLPTLPKPSHANKDHKS 307
QY 163 RFRYS-----YLLRELKAIKLVFEGKVTLKVKDIEAVPVVAVNTAVMLIGRLPLM 216
Db 308 FRIHKVYKSYIPLKVRKSFDFPD---TLMRKLLKGLSKWTGIDE-----VYRLKVM 360
QY 217 N-PKKVASWIEDKNILL-----YCTDIEFIGYRDIAGRMSVEGLLE 256
Db 361 HLOAKKLLLGISEETLTGLKERSVDEALLFAVRG-----NVDFLVE 405
QY 257 VIDEINSELCPSELKSHGRELRLYRTSSWADKSLRIWREDEGNARLNMLYN--RGEALFL 314
Db 406 MIRN-NSPELLWSTRTSSSTLFLAVEFRQEKV-----FNLLYGLDDRKYLLA 453
QY 315 AENSAR-----GWPLPERRL 330
Db 454 DKDSGNGVYLHLAGFPSPSKL 475

RESULT 9
ID Q9UYC0 PRELIMINARY; PRT; 394 AA.
AC Q9UYC0;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE HYPOTHETICAL 44.0 KDA PROTEIN.
GN PAB1309.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RT Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248288; CAB50492.1;
DR InterPro: IPR001696; Na_channel.
DR PRINTS: PRO0170; NACHANNEL.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 394 AA; 43951 MW; 11001889D1AB1165 CRC64;

Query Match 5.7%; Score 101; DB 1; Length 394;
Best Local Similarity 21.5%; Pred. No. 27;
Matches 53; Conservative 43; Mismatches 66; Indels 84; Gaps 13;
QY 7 HGNLOVAEIPKSPKVEKAYIPVETLIKKEPFGNITGYTLKFLPKDIIILVKGGIASD 66
Db 72 HGN-----IVLALSPIGLII-----GLGI-GYFLYY-----TGGWASG 104
QY 67 LIETIGSYTAIPLPLSLRVEAQVDRVKEELFEVSPKGFWLPALADPIIPAILKDN 126
Db 105 DVIILG-AYSALLPYVP-----DSAKYK-----PPSYSLPM-----NA 137
QY 127 YEYLFADAMLSAHLNSAIKPIKPLHLIKAQREKFRYISYLLRELKAIKLVFEGR 186
Db 138 FTILNLSLLIFPLILYSV-----VGLAVKGIKRELGVFREG- 176
QY 187 VTLKVKDIEAVPVVAVNTAVMLIGRLP-LMPKKVASWIEDKNILYGTDIEFIGYRDI 245
Db 177 ----IRNVVEVTIW--INFGAVLLAFVSLHMSIPKIISWILTFALILFSRFLVG--DV 228
QY 246 AGRMSV 251
Db 229 LGVLSI 234

RESULT 10
ID Q32450 PRELIMINARY; PRT; 653 AA.
AC Q32450;
DT 01-JAN-1998 (TremBLrel. 05, Created)
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE 4-ALPHA-GLUCANOTRANSFERASE.
OS Pyrococcus kodakaraensis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=69014;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1;
RA Tachibana Y., Fujiwara S., Takagi M., Imanaka T.;
RL J. Ferment. Bioeng. 83:540-548(1997).
DR EMBL: D87907; BAA22062.1;
KW Transferase.
SQ SEQUENCE 653 AA; 76664 MW; B0C4695613F29219 CRC64;

Query Match 5.7%; Score 100.5; DB 1; Length 653;
Best Local Similarity 19.3%; Pred. No. 63;
Matches 74; Conservative 67; Mismatches 127; Indels 115; Gaps 19;
QY 24 EKAYIPVETLIKKEP---FGLNITGYTLKFL-----PKDIIILVKGGIASDIEIGTS-Y 75
Db 28 ERSTRPFMETL-EYPMNKVAVHYSGPLLEWIRDNKPEHLDDLRLSLVKRGOLEIVVAGFY 86
QY 76 TAILPLPLSLRVEAQVDRVKE--ELFEVSPKGFWLP-LADPIIPAILKDNQYEVYLF 132
Db 87 EPVLASIP--KEDRIVQIEKLKEFARNLGYEARGWLTERTVWQPELVKSLRAAGIDYIV 144
QY 133 DEAMLSAHLNS-----AIKPIKPLHLIKAQREKFRYISYLLRELRL 176
Db 145 DDYHFMASGLSKDELFWPYTTEGGEVITVFPI-----DEKLRYL--IPFRPVD 191
QY 177 KAIKLVF-----EGKVTLKVKDIEAVPVVAVNTAVMLIGRLPLMNPKKVASWIEDK-- 228
Db 192 KTLLEYLHSDGDGSKVAVFHDGGEKFGWPG-----TYEWVYKGW 233
QY 229 -----NLLYGTDIE-----FIGYRDIAG--RMSVEGL-----LEVID 259
Db 234 LREPPDRVSSDERINLMLYSEVLRFRPRGLVYLPISYFEMSEWSLPAQAKLFVEVE 293
QY 260 ELNSELCPSELKSHGRELRLYRTSSWADKSLRIWREDEGNARLNMLYNRGLAFLAENS 319
Db 294 ELK-----KENKFDYRVFVRGGIWKNEFFKYPESNYHKKMLVSKA-----VRNPE 342
QY 320 ARGWPLPERRLDAFRAIYNDWRG 342
Db 343 AREFTLRACNDAY-----NHG 359

RESULT 11
ID Q9PKB1 PRELIMINARY; PRT; 690 AA.
AC Q9PKB1;
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN TC0556.
GN TC0556.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG.
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,

Db 118 REVNGNKTIIWKDPLKQSEV-----EALYELPEGEVIARYKDG-KPAAIKINKKI 170
QY 129 YL--FADAMLPES-AHLNSAIKPIKPLHLTKAOREKFRYISYLLRLKRAIKLVFEG 185
Db 171 YVGFPEDEVLNLIYHIVKTSNPLPYIL-----ETVFLTLASL-----MLTFQ- 216
QY 186 KVTLVKQDIEAVPVVAVNTAVMLIGRLPLANPKKVASWIEDKNILLYGTDIIEFIGYDI 245
Db 217 -ETLKKKFLLEISALASVK--VFILSRINLLDEKV-----LNDTTRREIYNY- 261
QY 246 AGRMVEGELLEVIDELNSELCPSELKHSRGRELY-----LRTSSWADKSLRWREDEGNA 299
Db 262 -----ILDNPGCHL-----RELSKNLNKPVSTLW-----HLRILEK-----A 294
QY 300 RLNNLYNMRGELAFLENSDARGWPLPERRLDFAFIY 337
Db 295 NLIKSKKEGNRIIYPADMDRDLPLLYLKNETQKSIF 332

RESULT 14
Q9AVF4 PRELIMINARY; PRT; 608 AA.
ID Q9AVF4
AC Q9AVF4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE MG-CHELTASE SUBUNIT CHLH (FRAGMENT).
GN MG-CH.
OS Amaranthus tricolor (Tampala).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Amaranthaceae; Amaranthus.
OX NCBI_TaxID=29722;
RN [1]
RP SEQUENCE FROM N.A.
RA Iwamoto K., Fukuda H., Sugiyama M.;
RT "Elimination of POR Expression correlates with Red Leaf Formation in
RT Amaranthus tricolor."
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050121; BAB41188.1;
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 608 AA; 67073 MW; 5D262E62BE2343CD CRC64;

Query Match 5.6%; Score 99; DB 10; Length 608;
Best Local Similarity 21.9%; Pred. No. 72;
Matches 79; Conservative 62; Mismatches 115; Indels 104; Gaps 19;
QY 13 AEIPKSE-PKVIKAVIPVIETLIKKEPFGNITGYTLKPLPKDIIIV-----KGGI 63
Db 111 SELPTAERDLVVRKIYAKIMEIESRLFPCCGLHVIGEPPTALEAVATLVNTAALDRPDGI 170
QY 64 AS-----DLIEIGTSTYAILPLPLSRVEAQVORV-----KEELPE 102
Db 171 SSILPSILAQTGVGRDIEDVYRGNDRGILKDVLELLRQITETSRDAITAFVERTNDKGVVN 230
QY 103 VSPK-----GFWLPPELADPTIIPAILKDNNGEYFLFADEAML--FSAHLSNAIKPIKPLHL 155
Db 231 VNDKLTSLILGFGLNE-----PWIQVLSNTKFRADREKLRVLFAPLGCLK-----L 277
QY 156 IKAQREKFRYISYLLRLRKAI-----KLVPFGKVTLLKVKDIEAVP 198
Db 278 VVADNE-----LGSLLKQALEGKYVEPGGDPINPKVLTGK-NIHALDPQAI 326
QY 199 VWVAVNTAVMLIGRLPLMPKVK--ASWIEDKNILLYGTDI-----IEFIGYR 243
Db 327 TTAAMQSAKVVVDR--LLERQADNGGKYPTVALVVLWGTDNKTGYESLAQVLMWIGVR 384
QY 244 DIA---GRMS-VEGLLEVIDELNSELCPSELKHSRG--RELYLRTSSWADKSLRWRE-DE 296
Db 385 PVADTFGRVKNKEPV--SLEELGRVTDVVVNCVSGFRDLFINQNNLLDRVAKMUTELDE 442

RESULT 15
O17474 PRELIMINARY; PRT; 742 AA.
ID O17474
AC O17474
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CYCLIC GMP-DEPENDENT PROTEIN KINASE.
GN HYCK.
OS Hydra oligactis (Hydra).
OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroids; Anthomedusae;
OC Hydridae; Hydra.
OX NCBI_TaxID=6088;
RN [1]
RP SEQUENCE FROM N.A.
RA Haeusler C.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
CC 1- SIMILARITY: TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; AF031931; AAC23588.1;
DR HSSP; P05132; IATP.
DR InterPro; IPR000595; CNMP_binding.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002052; N6_Mtase.
DR InterPro; IPR000961; Pkinase_C.
DR InterPro; IPR000215; Serpin.
DR InterPro; IPR002290; Ser_thr_kin_actsite.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00027; CNMP_binding; 2.
DR Pfam; PF00069; pkinase; 1.
DR SMART; SM00100; CNMP; 2.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM00133; S_TK_X; 1.
DR PROSITE; PS00888; CNMP_BINDING_1; 1.
DR PROSITE; PS00889; CNMP_BINDING_2; 2.
DR PROSITE; PS00442; CNMP_BINDING_3; 2.
DR PROSITE; PS00092; N6_MTASE; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00284; SERPIN; UNKNOWN_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 742 AA; 85345 MW; B105D8C09618D154 CRC64;
Query Match 5.6%; Score 99; DB 5; Length 742;
Best Local Similarity 21.7%; Pred. No. 96;
Matches 76; Conservative 60; Mismatches 133; Indels 82; Gaps 14;
QY 10 LOYAEIPKSEPKVIKAVIPVIETLIKKEPFGNITGYTLKF--LPKDIILVKG----- 61
Db 21 LDFARKLHYPLK--RQDLKKKDELIEEK--NKKLSGYEIDILNKKKEIRQLQSIVIEATR 76
QY 62 --GIASDLIEIIGTSTYAILPLPLSRVEAQVOR-----DRVKEELFVSPK-GFWLPPEL 113
Db 77 NTNNAVNDVIVEEEDNGILLSRHSLSQNHKPSRFIEVYVTRVNRKFRFAVSAESGKELQN 136
QY 114 ADPIIP-----AILKDNNGEYFLFADEAMLFSAHLSNAIKPIKPLHLKIAQREK 162
Db 137 DFPKIPKSTEVKEFTSQAVLANNFLKHL--ECQVKEIVLFMSOKSPKRGYIIEKEDMG 194
QY 163 RFRYSYLLRLKRAIKLVFEGKV-----TLVKVDIEAVPVW----- 200
Db 195 NALFVSYVGLLEISQGGKIL--GKPLRPGELFGELAILNCTRTASVKAIDDEVWYLER 252
QY 201 -----VAVNTAVM-----LIGRLPLANPKKVASWIEDKNILLYGTDIPI-- 240
Db 253 NVFOAVQKQTCIMRREHYNFLHSPVPFKNLPNDTLTKIVEIEE-----FYDEGEFIVR 308
QY 241 -GYRDIAGRMSEVGLLEVIDELNSELCPSELKHSRGRELYLRTSSWADKSLR 290

Thu Apr 4 16:34:58 2002

us-09-407-806a-4_1.rspt

Page 8

Db 309 EGERGDSFYILKQGLVKVLQMEGKDEPVEIRHLSQGEYFGEKALLGEDVVR 359

Search completed: April 4, 2002, 14:44:28
Job time: 252 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2002, 14:41:21 ; Search time 25.55 seconds
(without alignments)
1003.106 Million cell updates/sec

Title: US-09-407-806A-4
Perfect score: 346
Sequence: 1 LRLVFNHLYAEIPKSEP.....ERRLDAPRAIYNDWRNGCEP 346

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : A_Geneseq.ll01.*
- 1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	19	5.5	364	AAW34643	Thermotable alpha
2	8	2.3	466	AAW59165	L. lactis MG1316 g
3	8	2.3	624	AAW35671	Chlamydia pneumoni
4	7	2.0	47	AAG77652	Human colon cancer
5	7	2.0	50	AAW30440	Amino acid sequenc
6	7	2.0	83	AAW08113	Arabidopsis thalia
7	7	2.0	83	AAW40493	Arabidopsis thalia
8	7	2.0	108	AAW08112	Arabidopsis thalia
9	7	2.0	108	AAW40492	Arabidopsis thalia
10	7	2.0	113	AAW13155	Human secreted pro
11	7	2.0	116	AAW12375	Human 5' EST seque

12	7	2.0	138	21	AAW54427	Human pancreatic c
13	7	2.0	140	20	AAW37528	Amino acid sequenc
14	7	2.0	142	21	AAG40491	Arabidopsis thalia
15	7	2.0	143	21	AAG08111	Arabidopsis thalia
16	7	2.0	156	21	AAG04247	Arabidopsis thalia
17	7	2.0	178	20	AAW50041	Buchnera aphidicol
18	7	2.0	225	21	AAG20264	Arabidopsis thalia
19	7	2.0	225	22	AAW50345	Arabidopsis thalia
20	7	2.0	225	22	AAU04821	Arabidopsis thalia
21	7	2.0	268	21	AAG20263	Micromonospora eve
22	7	2.0	268	21	AAG50344	Arabidopsis thalia
23	7	2.0	286	21	AAW35951	Human 3-hydroxyacy
24	7	2.0	301	21	AAW87599	Human fatty acid b
25	7	2.0	306	22	AAG92081	C glutamicum prote
26	7	2.0	310	20	AAW13370	Amino acid sequenc
27	7	2.0	310	21	AAW27650	Human protein PRO2
28	7	2.0	310	22	AAW80238	Human PRO238 prote
29	7	2.0	325	21	AAW92513	Human OXRE-10. Ho
30	7	2.0	325	21	AAW73426	Human secreted pro
31	7	2.0	325	22	AAW38738	Human polypeptide
32	7	2.0	327	22	AAW40524	Human polypeptide
33	7	2.0	337	22	AAW42410	Human polypeptide
34	7	2.0	350	19	AAW74861	Human secreted pro
35	7	2.0	358	22	AAG92635	C glutamicum prote
36	7	2.0	382	21	AAG20282	Arabidopsis thalia
37	7	2.0	382	21	AAW50343	Arabidopsis thalia
38	7	2.0	460	20	AAW02536	Human ICAM-6 prote
39	7	2.0	471	22	AAW98380	Escherichia coli p
40	7	2.0	484	18	AAW21897	Glutamyl-tRNA synt
41	7	2.0	486	20	AAW14129	Human IL-1RD10 pro
42	7	2.0	696	21	AAW95238	Human krec2. Homo
43	7	2.0	867	20	AAW37095	Protein involved i
44	7	2.0	870	20	AAW35233	C. pneumoniae prot
45	7	2.0	999	21	AAW73362	HTRM clone 2674047

ALIGNMENTS

RESULT 1
AAW34643
ID AAW34643 standard; Protein; 364 AA.
XX
AC AAW34643;
XX
DT 27-MAR-1998 (first entry)
XX
DE Thermotable alpha-galactosidase AEDII12RA-alpha-gal-18GC.
XX
KW Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;
KW food processing; alpha glycoside hydrolysis; raffinose;
KW stachyose; verbascose; bean; flutulence; AEDII12RA-alpha-18GC.
XX
OS Thermococcus alcaliphilus strain AEDII12RA.
XX
FH Key Location/Qualifiers
FT Misc-difference 329 /note= "encoded by CTT"
FT
XX
PN WO9732974-A1.
XX
PD 12-SEP-1997.
XX
PF 05-FEB-1997; 97WO-US01452.
XX
PR 08-MAR-1996; 96US-0613220.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Murphy D, Reid J, Rudolph MJ;
XX
DR WPI; 1997-470541/43.
N-PSDB; AAT93753.

XX Nucleic acid encoding alpha-galactosidase from Thermococcus
PT alcaliphilus - used in food processing to hydrolyse
PT alpha-glycosides, e.g. raffinose
XX
PS Claim 1; Fig 1; 32pp; English.
XX This protein comprises AEDIII2RA-alpha-gal-186C, a claimed
CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
CC AEDIII2RA, a bacterium that shows optimum growth at 85 deg C and
CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
CC AAY93753) encoding the alpha-galactosidase; (2) a vector containing
CC the polynucleotide or homologous or complementary sequences; (2)
CC host cells containing the vector; (3) a process for producing the
CC alpha-galactosidase in transformed or transfected host cells; an
CC enzyme showing at least 70% identity to alpha-galactosidase and
CC comprising at least 30 amino acid residues of its sequence; and (4)
CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
CC in sugar beet processing (raffinose inhibits crystallisation of
CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose
CC and verbascone in beans and other gassy foods.
XX
SQ Sequence 364 AA;

Query Match 5.5%; Score 19; DB 18; Length 364;
-Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LRALVFHGNLQYAEIPKSE 19
Db 1 lralvfghnldyaelpkse 19
|||||

RESULT 2
AAW59165
ID AAW59165 standard; Protein; 466 AA.
XX
AC AAW59165;
XX
DT 13-AUG-1998 (first entry)
XX
DE L. lactis MG1316 gadB protein from PstI/EcoRI fragment.
XX
XX Salt-inducible promoter; lactic acid; food industry; food-grade inducer;
KW fermentation processes; cheese production; rnhB; gadB; gadC; orfX; rggL.
XX
XX Lactococcus lactis.
OS
XX
PN WO9810080-A1.
XX
PD 12-MAR-1998.
XX
XX 20-AUG-1997; 97WO-EP04755.
XX
XX 13-MAR-1997; 97EP-0200744.
PR 05-SEP-1996; 96EP-0202444.
XX
XX (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
XX
XX Kok J, Ledebor AM, Sanders JW, Venema G;
XX
XX WPI: 1998-193629/17.
DR N-PSDB; AAV11905.
XX
XX Salt-inducible promoter - derived from lactic acid bacteria, used
PT for the production of polypeptides in food
XX
XX Example 6; Fig 29; 111pp; English.
XX
XX This protein sequence is the gadB protein found in a PstI-EcoRI fragment

CC from the NS3 locus of the lactic acid bacterium Lactococcus lactis. This
CC fragment also contains the C-terminus of rnhB, full length gadR protein
CC (also known as rggL), gadC protein (also known as orfX) and the
CC C-terminus of an unknown reading frame which reads in the reverse
CC orientation. This fragment is also found to act as a salt-inducible
CC promoter (SIP). Using this SIP, salt can be used as a food-grade inducer
CC in food fermentation processes, e.g. in the production of cheese,
CC dressings, water-containing spreads, sausages, or sour dough.
XX
SQ Sequence 466 AA;

Query Match 2.3%; Score 8; DB 19; Length 466;
-Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 295 DEGNARLN 302
Db 52 degnarln 59
|||||

RESULT 3
AAY35671
ID AAY35671 standard; Protein; 624 AA.
XX
AC AAY35671;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae transmembrane protein sequence.
XX
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
XX Chlamydia pneumoniae.
OS
XX
PN WO9927105-A2.
XX
PD 03-JUN-1999.
XX
XX 20-NOV-1998; 98WO-IB01890.
XX
XX 04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
XX (GEST) GENSET.
PA
XX
XX Griffais R;
PI
XX
XX WPI: 1999-357842/30.
DR
XX
XX Genome sequence of Chlamydia pneumoniae
PT
XX
XX Page 1387-1388; Disclosure; 1912pp; English.
XX
XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC C. pneumoniae is thought to be a contributing factor in heart
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
XX
SQ Sequence 624 AA;

Query Match 2.3%; Score 8; DB 20; Length 624;
-Best Local Similarity 100.0%; Pred. No. 25;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 PAILKDNQ 126
|||||||
Db 503 pailkdnq 510

RESULT 4

AAG77652 AAG77652 standard; Protein: 47 AA.

XX AC AAG77652;

XX DT 03-SEP-2001 (first entry)

XX Human colon cancer antigen protein SEQ ID NO:8418.

XX Human: colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.

XX OS Homo sapiens.

XX PN WO200122920-A2.

XX PD 05-APR-2001.

XX PF 28-SEP-2000; 2000WO-US26524.

XX PR 29-SEP-1999; 99US-0157137.

XX PR 03-NOV-1999; 99US-0163280.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX WPI; 2001-235357/24.

XX DR N-PSDB; AAH37059.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 11; Page 9686; 9803pp; English.

XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAB7789 represent sequences used in the exemplification of the
XX present invention.

XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 47 AA;

Query Match

Best Local Similarity 2.0%; Score 7; DB 22; Length 47;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

-QY 165 RYISYLL 171

Db 24 ryisyll 30

RESULT 5

AAB30440 AAB30440 standard; peptide; 50 AA.

XX AC AAB30440;

XX DT 06-MAR-2001 (first entry)

XX Amino acid sequence of a fragment of a galectin family member.

XX Galectin; Galectin-3; cell growth; tumour; metastasis; cirrhotic liver;
XX hepatocellular carcinoma; liver neoplasm.

XX OS Gallus sp.

XX PN WO200062076-A1.

XX PD 19-OCT-2000.

XX PF 29-MAR-2000; 2000WO-US08561.

XX PR 13-APR-1999; 99US-0129111.

XX PA (HSUD/) HSU D K.

XX PI Hsu DK, Liu F, Dowling CA;

XX WPI; 2000-679520/66.

XX Cirrhotic liver and hepatocellular carcinoma marker galectin-3, useful
XX in diagnostic, prognostic and therapeutic methods for treating and
XX preventing hepatocellular carcinoma -

XX Example 3; Page 35; 49pp; English.

XX AAB30440-85 represent peptides derived from galectin family members.
XX Galectin-3 has been associated with assorted processes such as cell
XX growth, tumour transformation, and metastasis. It is composed of two
XX domains: a carbohydrate-binding region on the carboxy terminal which
XX binds saccharides with tandem repeats of nine amino acids. The expression
XX of galectin-3 is induced in cirrhotic liver and hepatocellular
XX carcinoma, and so is useful as a marker of cirrhotic liver and
XX hepatocellular carcinoma. Galectin-3 is used in diagnostic, prognostic
XX and therapeutics for the treatment and prevention of hepatocellular
XX carcinoma. It is useful for determining whether an individual is
XX suffering from hepatocellular carcinoma or is likely to suffer from
XX hepatocellular carcinoma or any neoplasm of liver origin or cirrhosis of
XX the liver in the future.

XX Sequence 50 AA;

Query Match

Best Local Similarity 2.0%; Score 7; DB 21; Length 50;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 79 LPILPLS 85

Db 15 lpilpls 21

RESULT 6

AAG08113 AAG08113 standard; Protein: 83 AA.

XX AC AAG08113;

XX DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 5523.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140931.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.


```
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 07-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 83;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 KVTLKVK 192
Db 2 kvtlkvk 8
|||||

RESULT 8
AAG08112
ID AAG08112 standard; Protein; 108 AA.
XX
AC AAG08112;
XX
XX 17-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 5522.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR
```

PR	05-MAR-1999;	99US-0123180.	PR	14-JUL-1999;	99US-0143624.
PR	09-MAR-1999;	99US-0123548.	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	99US-0125788.	PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	99US-0126264.	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	99US-0126785.	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	99US-0127462.	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	99US-0128234.	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	99US-0128845.	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	99US-0130077.	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	23-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144884.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0144814.
PR	04-MAY-1999;	99US-0132407.	PR	21-JUL-1999;	99US-0145086.
PR	05-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145088.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145087.
PR	07-MAY-1999;	99US-0132863.	PR	22-JUL-1999;	99US-0145089.
PR	11-MAY-1999;	99US-0134218.	PR	22-JUL-1999;	99US-0145192.
PR	14-MAY-1999;	99US-0134219.	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	99US-0134221.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134370.	PR	23-JUL-1999;	99US-0145224.
PR	18-MAY-1999;	99US-0134768.	PR	26-JUL-1999;	99US-0145276.
PR	19-MAY-1999;	99US-0134941.	PR	27-JUL-1999;	99US-0145913.
PR	20-MAY-1999;	99US-0135124.	PR	27-JUL-1999;	99US-0145918.
PR	21-MAY-1999;	99US-0135353.	PR	27-JUL-1999;	99US-0145919.
PR	24-MAY-1999;	99US-0135629.	PR	28-JUL-1999;	99US-0145951.
PR	25-MAY-1999;	99US-0136021.	PR	02-AUG-1999;	99US-0146386.
PR	27-MAY-1999;	99US-0136392.	PR	02-AUG-1999;	99US-0146388.
PR	28-MAY-1999;	99US-0136782.	PR	03-AUG-1999;	99US-0147038.
PR	01-JUN-1999;	99US-0137222.	PR	04-AUG-1999;	99US-0147204.
PR	03-JUN-1999;	99US-0137528.	PR	04-AUG-1999;	99US-0147302.
PR	04-JUN-1999;	99US-0137502.	PR	05-AUG-1999;	99US-0147192.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147260.
PR	08-JUN-1999;	99US-0138094.	PR	06-AUG-1999;	99US-0147416.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138847.	PR	09-AUG-1999;	99US-0147935.
PR	14-JUN-1999;	99US-0139119.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148341.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	99US-0139455.	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	99US-0139456.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139457.	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139460.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139462.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139463.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139750.	PR	25-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139750.	PR	26-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151066.
PR	21-JUN-1999;	99US-0139817.	PR	27-AUG-1999;	99US-0151080.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151303.
PR	23-JUN-1999;	99US-0140353.	PR	31-AUG-1999;	99US-0151438.
PR	23-JUN-1999;	99US-0140354.	PR	01-SEP-1999;	99US-0151930.
PR	24-JUN-1999;	99US-0140695.	PR	07-SEP-1999;	99US-0152363.
PR	28-JUN-1999;	99US-0140823.	PR	10-SEP-1999;	99US-0153070.
PR	29-JUN-1999;	99US-0140991.	PR	13-SEP-1999;	99US-0153758.
PR	01-JUL-1999;	99US-0141287.	PR	15-SEP-1999;	99US-0154018.
PR	01-JUL-1999;	99US-0141842.	PR	16-SEP-1999;	99US-0154039.
PR	01-JUL-1999;	99US-0142154.	PR	20-SEP-1999;	99US-0154779.
PR	02-JUL-1999;	99US-0142055.	PR	22-SEP-1999;	99US-0155139.
PR	06-JUL-1999;	99US-0142390.	PR	23-SEP-1999;	99US-0155486.
PR	08-JUL-1999;	99US-0142803.	PR	24-SEP-1999;	99US-0155565.
PR	09-JUL-1999;	99US-0142920.	PR	28-SEP-1999;	99US-0156458.
PR	12-JUL-1999;	99US-0142977.	PR	28-SEP-1999;	99US-0156596.
PR	13-JUL-1999;	99US-0143542.	PR	04-OCT-1999;	99US-0157117.

```
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 108;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 KVTLLKVK 192
Db 27 kvllkvvk 33

RESULT 9
AAG40492
ID AAG40492 standard; Protein; 108 AA.
AC AAG40492;
XX
XX
DT 18-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 50249.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX KW termination sequence.
XX
XX OS Arabidopsis thaliana.
XX
XX PN EPI033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX PR 25-FEB-1999; 99US-0121825.
XX PR 05-MAR-1999; 99US-0123180.
XX PR 09-MAR-1999; 99US-0123348.
XX PR 23-MAR-1999; 99US-0125788.
XX PR 25-MAR-1999; 99US-0126264.
XX PR 29-MAR-1999; 99US-0126785.
XX PR 01-APR-1999; 99US-0127462.
XX PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
```

pt New isolated brain-derived nucleic acids - used to develop products which may have cytokine, immune, regulatory, haematopoiesis regulating, anti-inflammatory or tumour inhibition activity

XX Claim 34; Page 541; 577pp; English.

XX AAX51787 to AAX52019 represent 5' expressed sequence tags (ESTs) for

CC human secreted proteins, and encode the proteins given in AAY12987 to

CC AAY13219, respectively. The proteins given represent the signal peptide

CC and an N-terminal fragment of a secreted protein. The nucleic acid

CC sequences can be used for producing secreted human gene products. They

CC can also be used to develop products for diagnosis and therapy. The

CC proteins obtained may have cytokine activity, cell

CC proliferation/differentiation activity, haematopoiesis regulating

CC activity, tissue growth regulating activity, reproductively hormone

CC regulating activity, chemotactic/chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity or other activities. The products

CC can be used in forensic, gene therapy and chromosome mapping procedures.

CC The sequences can also be used for obtaining corresponding promoter

CC sequences. The nucleic acids encoding the signal peptide can be used for

CC directing extracellular secretion of a polypeptide or the insertion of a

CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 113 AA;

SQ

Query Match 2.0%; Score 7; DB 20; Length 113;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 TAILPL 82

Db 27 tailpll 33

RESULT 11

AAY12375

ID AAY12375 standard; Protein; 116 AA.

AC AAY12375;

XX

XX 17-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID NO:406.

XX Human; secreted protein; EST: expressed sequence tag; diagnosis;

KW forensic; gene therapy; chromosome mapping; signal peptide;

KW upstream regulatory sequence; cytokine activity; cell proliferation;

KW differentiation; haematopoiesis regulation; tissue growth regulation;

KW reproductively hormone regulation; chemotactic; chemokinetic; haemostatic;

KW thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

OS

XX WO9506548-A2.

PN

XX 11-FEB-1999.

PD

XX 31-JUL-1998; 98WO-IB01222.

PF

XX 01-AUG-1997; 97US-0905135.

PR

XX (GIST) GENSET.

PA

XX Duclert A, Dumas Milne Edwards J, Lacroix B;

PI WPI; 1999-153778/13.

DR N-PSDB; AAX41208.

DR

XX New nucleic acids encoding human secreted proteins - obtained from

PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,

PT kidney, lung, umbilical cord, placenta and colon tissue

XX

PS Claim 27; Page 727-728; 824pp; English.

XX

CC AAX41094 to AAX41347 represent 5' expressed sequence tags (ESTs) for

CC human secreted proteins, and encode the proteins given in AAY12261 to

CC AAY12514, respectively. The proteins given represent the signal peptide

CC and an N-terminal fragment of a secreted protein. The nucleic acid

CC sequences can be used for producing secreted human gene products. They

CC can also be used to develop products for diagnosis and therapy. The

CC proteins obtained may have cytokine activity, cell

CC proliferation/differentiation activity, haematopoiesis regulating

CC activity, tissue growth regulating activity, reproductively hormone

CC regulating activity, chemotactic/chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity or other activities. The products

CC can be used in forensic, gene therapy and chromosome mapping procedures.

CC The sequences can also be used for obtaining corresponding promoter

CC sequences. The nucleic acids encoding the signal peptide can be used for

CC directing extracellular secretion of a polypeptide or the insertion of a

CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 116 AA;

SQ

Query Match 2.0%; Score 7; DB 20; Length 116;

Best Local Similarity 100.0%; Pred. No. 52;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 76 TAILPL 82

Db 7 tailpll 13

RESULT 12

AAB54427

ID AAB54427 standard; Protein; 138 AA.

AC AAB54427;

XX

XX 09-MAR-2001 (first entry)

DT Human pancreatic cancer antigen protein sequence SEQ ID NO:879.

DE

XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

KW detection; diagnosis; identification; cytostatic; neuroprotective;

KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;

KW anti-inflammatory; cardiant; gene therapy; chromosome mapping;

KW linkage analysis; tissue identification; tissue typing; forensic;

KW neural; immune system; muscular; reproductively; gastrointestinal;

KW pulmonary; cardiovascular; renal; proliferative.

XX Homo sapiens.

OS

XX WO200055320-A1.

PN

XX 21-SEP-2000.

PD

XX 08-MAR-2000; 2000WO-US05989.

PF

XX 12-MAR-1999; 99US-0124270.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX Rosen CA, Ruben SM;

PI WPI; 2000-579444/54.

XX N-PSDB; AAC99192.

DR

XX New nucleic acid that is a pancreatic cancer antigen for preventing,

PT treating, or ameliorating a medical condition, particular pancreatic

PT cancer, or for use in assays for diagnosing a pathological condition -

PS Claim 11; Page 1339; 1379pp; English.

XX

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated

CC proteins, called pancreatic cancer antigens, given in AAB54008 to

CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.

XX Sequence 138 AA;

Query Match 2.0%; Score 7; DB 21; Length 138;
 Best Local Similarity 100.0%; Pred. No. 61;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 308 RGEAFL 314
 Db 4 rgeafl 10
 |||||

RESULT 13

AAY37528
 ID AAY37528 standard; Protein; 140 AA.

XX AC AAY37528;

XX DT 07-OCT-1999 (first entry)

XX DE Amino acid sequence of a Chlamydia trachomatis protein.

XX KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumonia; venereal lymphogranulomatosis.

XX OS Chlamydia trachomatis.

XX PN WO9928475-A2.

XX PD 10-JUN-1999.

XX PF 27-NOV-1998; 98WO-1801939.

XX PR 04-NOV-1998; 98US-0107077.

XX PR 28-NOV-1997; 97FR-0015041.

XX PR 17-DEC-1997; 97FR-0016034.

XX PA (GEST) GENSET.

XX PI Griffiths R;

XX DR WPI; 1999-371125/31.

XX PT Genome sequence of Chlamydia trachomatis

XX PS Disclosure; Page 1196; 1755pp; English.

XX PS AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

XX CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as

XX CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC Bartholinitis, Bartholinitis; pneumonia; venereal lymphogranulomatosis.
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.

XX Sequence 140 AA;

Query Match 2.0%; Score 7; DB 20; Length 140;
 Best Local Similarity 100.0%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 KEELFEV 103

Db 102 keelfev 108
 |||||

RESULT 14

AAG40491

ID AAG40491 standard; Protein; 142 AA.

XX AC AAG40491;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 50248.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

XX PR 05-MAR-1999; 99US-0123180.

XX PR 09-MAR-1999; 99US-0123548.

XX PR 23-MAR-1999; 99US-0125788.

XX PR 25-MAR-1999; 99US-0126264.

XX PR 29-MAR-1999; 99US-0126785.

XX PR 01-APR-1999; 99US-0127462.

XX PR 06-APR-1999; 99US-0128234.

XX PR 08-APR-1999; 99US-0128714.

XX PR 16-APR-1999; 99US-0129845.

XX PR 19-APR-1999; 99US-0130077.

XX PR 21-APR-1999; 99US-0130449.

XX PR 23-APR-1999; 99US-0130510.

XX PR 28-APR-1999; 99US-0130891.

XX PR 28-APR-1999; 99US-0131449.

XX PR 30-APR-1999; 99US-0132048.

XX PR 30-APR-1999; 99US-0132407.

XX PR 04-MAY-1999; 99US-0132484.

XX PR 05-MAY-1999; 99US-0132485.

XX PR 06-MAY-1999; 99US-0132486.

XX PR 06-MAY-1999; 99US-0132487.

XX PR 07-MAY-1999; 99US-0132863.

XX PR 11-MAY-1999; 99US-0134256.

XX PR 14-MAY-1999; 99US-0134218.

XX PR 14-MAY-1999; 99US-0134219.

XX PR 14-MAY-1999; 99US-0134221.

XX PR 14-MAY-1999; 99US-0134370.

XX PR 18-MAY-1999; 99US-0134768.

XX PR 19-MAY-1999; 99US-0134941.

XX PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.

```
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      2.0%; Score 7; DB 21; Length 142;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 KVTIKVK 192
   |||||
Db 61 kvtkvk 67

RESULT 15
AAG08111
ID AAG08111 standard; Protein; 143 AA.
XX
AC AAG08111;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 5521.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137322.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
```

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 2.0%; Score 7; DB 21; Length 143;
Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 186 KVTLKVK 192
Db 62 kvtlkvk 68

Search completed: April 4, 2002, 14:45:30
Job time: 249 sec
